



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY

A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

META-HEURISTICKÉ ALGORITMY PRO VÝBĚR PŘÍZNAKŮ V KLASIFIKACI SRDEČNÍCH ONEMOCNĚNÍ

META-HEURISTIC ALGORITHMS FOR FEATURE SELECTION IN CLASSIFICATION OF HEART-RELATED DISEASES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Tereza Švestková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Martin Mézl, Ph.D.

BRNO 2024

Diplomová práce

magisterský navazující studijní program **Biomedicínské inženýrství a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Bc. Tereza Švestková

ID: 221532

Ročník: 2

Akademický rok: 2023/24

NÁZEV TÉMATU:

Meta-heuristické algoritmy pro výběr příznaků v klasifikaci srdečních onemocnění

POKYNY PRO VYPRACOVÁNÍ:

1) Seznamte se s postupy pro optimalizaci výběru příznaků v rámci úlohy klasifikace. 2) Proveďte rešerši meta-heuristických algoritmů, které by šly využít pro tuto optimalizaci. 3) Vyberte vhodnou klasifikační úlohu s dostupnými trénovacími daty a proveďte optimalizaci výběru příznaků standardními postupy (např. PCA) a vybraným meta-heuristickým algoritmem. 4) Práci rozšiřte o vhodnou úlohu z oblasti klasifikace srdečních onemocnění, pro kterou proveďte optimalizaci výběru příznaků. 5) Implementujte další algoritmy z oblasti meta-heuristických algoritmů a optimalizaci výběru příznaků zopakujte. 6) Proveďte diskuzi výsledků, vhodnosti použití jednotlivých metod a nastavení parametrů metod.

DOPORUČENÁ LITERATURA:

[1] AY, Şevket; EKINCI, Ekin a GARIP, Zeynep, 2023. A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. Online. The Journal of Supercomputing. Roč. 79, č. 11, s. 11797-11826. ISSN 0920-8542. Dostupné z: <https://doi.org/10.1007/s11227-023-05132-3>. [cit. 2023-09-05].

[2] ODILI, Julius Beneoluchi; AMBAR, Radzi; WAHAB, Mohd Helmy Abd, 2018. A Critical Review of Major Nature-Inspired Optimization Algorithms. Online. The Eurasia Proceedings of Science Technology Engineering and Mathematics. Roč. 2, s. 376-394. Dostupné z: <http://www.epstem.net/en/pub/issue/38904/455970>. [cit. 2023-09-05].

Termín zadání: 5.2.2024

Termín odevzdání: 22.5.2024

Vedoucí práce: Ing. Martin Mézl, Ph.D.

prof. Ing. Valentine Provazník, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato práce je věnována postupům při výběru relevantních příznaků pro klasifikační úlohy týkající se srdečních onemocnění. Optimální výběr příznaků je klíčovým faktorem pro správnou funkčnost klasifikačních modelů a v případě medicíny pro zlepšení diagnostiky. Teoretická část pojednává o obecné klasifikační úloze ve strojovém učení. Dále jsou podrobněji popsány některé klasické postupy i novější meta-heuristické algoritmy pro efektivní selekci příznaků. Praktická část se věnuje aplikaci některých popsaných algoritmů na datové soubory související se srdečním onemocněním. Na základě ověření validity výsledku klasifikačního modelu podle vybraných příznaků běžnými postupy i evolučními algoritmy jsou diskutovány výhody a přínosy upřednostnění meta-heuristických algoritmů.

KLÍČOVÁ SLOVA

Strojové učení, meta-heuristické algoritmy, klasifikace, příznaky, srdeční onemocnění

ABSTRACT

This thesis is devoted to the features selection for classification tasks related to heart disease. The optimal features selection is a key factor for the correct functionality of classification models and, in the case of medicine, for the improvement of diagnostics. The theoretical part discusses the general classification task in machine learning. Furthermore, some classic procedures as well as newer meta-heuristic algorithms for efficient feature selection are described in more detail. The practical part is devoted to the application of some of the described algorithms to data sets related to heart disease. The advantages and benefits of prioritizing meta-heuristic algorithms are discussed based on the verification of the validity of the result of the classification model according to selected symptoms of common procedures and evolutionary algorithms.

KEYWORDS

Machine learning, meta-heuristic algorithms, classification, features, heart disease

ŠVESTKOVÁ, Tereza. *Meta-heuristické algoritmy pro výběr příznaků v klasifikaci srdečních onemocnění*. Diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2024. Vedoucí práce: Ing. Martin Mézl, Ph.D.

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Bc. Tereza Švestková
VUT ID autora:	221532
Typ práce:	Diplomová práce
Akademický rok:	2023/24
Téma závěrečné práce:	Meta-heuristické algoritmy pro výběr příznaků v klasifikaci srdečních onemocnění

Prohlašuji, že svou závěrečnou práci jsem vypracovala samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autorky*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu diplomové práce panu Ing. Martinu Mězlovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	12
1 Optimalizace výběru příznaků pro obecnou klasifikační úlohu	13
1.1 Předzpracování dat	13
1.1.1 Normalizace	14
1.1.2 Standardizace	14
1.2 Klasifikační metody	15
1.2.1 Logistická regrese	15
1.2.2 Algoritmus K-nejbližších sousedů	15
1.2.3 Náhodná doménová struktura	16
1.3 Vyhodnocení správnosti predikce	16
1.3.1 Přesnost	16
1.3.2 Senzitivita	17
1.3.3 Specificita	17
1.3.4 F1 - skóre	17
2 Běžné metody pro výběr příznaků	18
2.1 Filtrační metody	18
2.1.1 Pearsonův korelační koeficient	18
2.1.2 Chí kvadrát test	18
2.2 Obálkové metody	18
2.2.1 Dopředné prohledávání	19
2.2.2 Zpětná eliminace	19
2.3 Redukce dimenzionality	19
2.3.1 PCA	19
3 Meta-heuristické algoritmy	20
3.1 Metaheuristické algoritmy inspirované inteligencí roje	21
3.1.1 Algoritmus podle kukaček	21
3.1.2 Algoritmus podle jestřábů Harrisových	21
3.1.3 Algoritmus podle velryb	24
3.1.4 Optimalizace podle papouška	25
3.1.5 Algoritmus opylování květin	26
3.1.6 Algoritmus umělého včelstva	27
3.1.7 Algoritmus mravenčí kolonie	28
3.1.8 Algoritmus světlušky	29
3.2 Metaheuristické algoritmy inspirované evolucí	30

3.2.1	Genetický algoritmus	30
3.2.2	Diferenciální evoluce	32
3.3	Metaheuristické algoritmy inspirované fyzikou	33
3.3.1	Simulované žíhání	33
4	Datasety a úprava dat	34
4.1	Kleveladský dataset srdečního onemocnění	34
4.1.1	Rozložení a úprava dat	34
4.2	Dataset srdečního selhání	36
4.2.1	Rozložení a úprava dat	36
4.3	Indický dataset srdečního onemocnění	37
4.3.1	Rozložení a úprava dat	37
5	Selekce příznaků pomocí klasického postupu - PCA	39
6	Selekce příznaků meta-heuristickými algoritmy	41
6.1	Selekce příznaků algoritmem papouška	41
6.2	Selekce příznaků algoritmem jestřábů	43
6.3	Selekce příznaků optimalizací opylování květin	44
6.4	Selekce příznaků pomocí algoritmu kukaček	46
6.5	Selekce příznaků optimalizací velryb	47
7	Srovnání klasifikace na základě vybraných příznaků všemi přístupy	49
	Závěr	59
	Literatura	60
	Seznam symbolů a zkratk	70
	Seznam příloh	71
A	Klevelandský dataset	72
B	Dataset srdečního selhání	73
C	Indický dataset	74
D	Pseudokódy použitých datasetů	75
E	Souhrnné výsledky pro vyzkoušené nastavení parametrů	78

Seznam obrázků

3.1	Rozdělení meta-heuristických algoritmů, převzato z [34]	20
3.2	Vývojový diagram genetického algoritmu, převzato z [63]	30
4.1	Histogram věkového rozložení	35
4.2	Histogram maximální tepové frekvence [tepy za minutu]	35
4.3	Histogram hodnoty kreatin-fosfokinázy v krvi [mcg/l]	36
4.4	Histogram kreatininu v séru [mg/dl]	37
4.5	Histogram maximální tepové frekvence [tepy za minutu]	38
4.6	Histogram klidového systolického krevního tlaku [mmHg]	38
5.1	Graf vývoje kumulativního součinu v závislosti na hlavních komponentách pro dataset srdečního selhání	39
6.1	Graf závislosti přesnosti predikce na parametru alfa u datasetu srdečního selhání	42
6.2	Graf závislosti přesnosti predikce na počtu jedinců pro Klevelandský dataset s parametrem alfa = 0,5	42
6.3	Graf závislosti přesnosti predikce na parametrem alfa pro dataset srdečního selhání	43
6.4	Graf závislosti přesnosti predikce na počtu jedinců s parametru alfa = 1,5 u Indického datasetu	44
6.5	Graf závislosti přesnosti predikce na parametru alfa pro Klevelandský dataset	45
6.6	Graf závislosti přesnosti predikce na počtu jedinců s alfa = 0,5 pro dataset srdečního selhání	45
6.7	Graf závislosti F1- skóre na nastavení parametru alfa při $p_a = 0,1$ pro Indický dataset	46
6.8	Graf závislosti F1- skóre na parametru p_a při konstantním alfa = 1,5 pro Klevelandský dataset	47
6.9	Graf závislosti F1- skóre na počtu jedinců při konstantní b pro Klevelandský dataset	48
6.10	Graf závislosti F1- skóre na změně parametru b u Indického datasetu	48
7.1	Graf závislosti výpočetního času na počtu jedinců při 100 iterací pro Klevelandský dataset	50
7.2	Graf závislosti F1- skóre na počtu jedinců a parametrech u CSO pro Klevelandský dataset	52
7.3	Graf závislosti F1- skóre na počtu jedinců a parametrech u FPA pro Klevelandský dataset	52
7.4	Graf závislosti F1- skóre na počtu jedinců a parametrech u HHO pro Klevelandský dataset	53

7.5	Graf závislosti F1- skóre na počtu jedinců a parametrech u PO pro Klevelandský dataset	53
7.6	Graf závislosti F1- skóre na počtu jedinců a parametrech u WOA pro Klevelandský dataset	54
7.7	Graf zastoupení počtu vybrání jednotlivých příznaků všemi algoritmy pro Klevelandský dataset	55
7.8	Graf zastoupení počtu vybrání jednotlivých příznaků všemi algoritmy pro dataset srdečního selhání	55
7.9	Graf zastoupení počtu vybrání jednotlivých příznaků všemi algoritmy pro Indický dataset	56

Seznam tabulek

5.1	Tabulka zastoupení příznaků, pro dané komponenty	40
7.1	Tabulka přesnosti klasifikace na základě všech příznaků	50
7.2	Tabulka přesnosti klasifikace na základě PCA	50
7.3	Tabulka nejlepších algoritmů na základě F1- skóre	51
7.4	Souhrnná tabulka srovnání výsledků studie a této práce na základě průměrných hodnot pro všechna nastavení daných algoritmů	57
7.5	Tabulka srovnání výsledků dostupných studií a této práce na základě hodnot F1- skóre	58
A.1	Tabulka příznaků s rozsahem hodnot pro Klevelandský dataset	72
B.1	Tabulka příznaků s rozsahem hodnot pro dataset srdečního selhání	73
C.1	Tabulka příznaků s rozsahem hodnot pro Indický dataset	74
E.1	Výsledky PO pro Klevelandský dataset	78
E.2	Výsledky PO pro Dataset srdečního selhání	79
E.3	Výsledky PO pro Indický dataset	80
E.4	Výsledky HHO pro Klevelandský dataset	81
E.5	Výsledky HHO pro Dataset srdečního selhání	82
E.6	Výsledky HHO pro Indický dataset	83
E.7	Výsledky FPA pro Klevelandský dataset	84
E.8	Výsledky FPA pro Dataset srdečního selhání	85
E.9	Výsledky FPA pro Indický dataset	86
E.10	Výsledky CSO pro Klevelandský dataset	87
E.11	Výsledky CSO pro Klevelandský dataset	88
E.12	Výsledky CSO pro Dataset srdečního selhání	89
E.13	Výsledky CSO pro Dataset srdečního selhání	90
E.14	Výsledky CSO pro Indický dataset	91
E.15	Výsledky CSO pro Indický dataset	92
E.16	Výsledky WOA pro Klevelandský dataset	93
E.17	Výsledky WOA pro Dataset srdečního selhání	94
E.18	Výsledky WOA pro Indický dataset	95

Úvod

Meta-heuristické algoritmy představují možnou budoucnost v optimální selekci příznaků v oblasti strojového učení. V kontextu srdečního selhání mohou pomoci vybrat relevantní příznaky pro predikci výskytu onemocnění. Jelikož srdeční onemocnění má často fatální následky, je vhodné mít informace o rizikových faktorech a snažit se rizika zmírnit včas.

Cílem této práce je seznámit čtenáře s klasickými postupy pro výběr příznaků i s relativně novými meta-heuristickými algoritmy inspirovanými přírodou. V teoretické části jsou shrnuty postupy pro vhodnou úpravu dat, základní algoritmy pro výběr příznaků a nejběžnější klasifikační modely.

V praktické části je aplikována analýza hlavních komponent a vhodné meta-heuristické algoritmy na datasey srdečních onemocnění pro relevantní výběr příznaků. Na základě tohoto výběru je naučen klasifikační model, který predikuje výskyt srdečního onemocnění. Cílem práce je srovnat přesnost výběru příznaků různými přístupy pro správnou klasifikaci a verifikovat tak možné výhody evolučních algoritmů pro využití nejen v medicíně.

1 Optimalizace výběru příznaků pro obecnou klasifikační úlohu

Optimalizace výběru příznaků je důležitou součástí klasifikačních modelů ve strojovém učení, dále jen ML. Klasifikační úlohy se využívají pro analýzu dat v různých odvětvích pro predikci určitých skupin na základě společných vlastností. Vhodným výběrem příznaků se snažíme zvýšit efektivitu klasifikace za použití menšího objemu dat. Odstraněním zbytečných příznaků lze dosáhnout rychlejšího běhu programu, dochází k redukci dimenzionality a snažší interpretaci výsledků. Mezi základní metody výběru příznaků patří filtrační metody, obálkové metody či analýza hlavních komponent, dále jen PCA. Mezi modernější algoritmy, které by potlačovaly nevýhody klasických metod, patří výběr příznaků založen na meta-heuristice. Meta-heuristika je založena na stochastice a funguje na principu "pokus- omyl", kde cílem je najít řešení v krátkém čase s využitím i neúplných informací. Možným využitím těchto algoritmů se zabývá studie [1] zaměřená na srdeční onemocnění. Pro všechny metody výběru příznaků je základním pilířem data pro danou klasifikační úlohu vhodně předzpracovat, což zajistí správnou validitu výsledků a nedojde k zanášení chyb do predikčního modelu. Po aplikaci výběru příznaků na předzpracovaná data dochází k predikci a vyhodnocení její přesnosti. Určité postupy pro obecnou klasifikační úlohu jsou popsány v následujících kapitolách. [2], [31]

1.1 Předzpracování dat

Důležitost předzpracování dat již byla zmíněna výše. Často se v datasetu objevují chybějící hodnoty, typicky využívaným řešením je nahradit nezastoupené hodnoty mediánem dané skupiny nebo úplné vyřazení vzorků, pokud skupina s neuvedenými daty není markantně zastoupena. Je však třeba si uvědomit, že ztrácíme informace, které mohou ovlivnit výsledek modelu. Mnohdy se vyskytují i odlehlé hodnoty, které mohou zapříčinit zcela negausovské rozložení dat, což může způsobit špatnou interpretaci výsledků, přílišnou adaptaci modelu či chybné shlukování dat. Běžným přístupem je tyto data nahradit mediánem tohoto příznaku z dané skupiny s podobnými parametry nebo opět data odstranit. Kategoriální proměnné zadané kvalitativně lze pro lepší přehlednost v datasetu převést na číselné hodnoty například (muž - 0, žena - 1). Kvantitativně vyjádřené příznaky lze škálovat pomocí normalizace či standardizace, upravit tak jejich rozptyl a zajistit srovnatelné vlivy na model. [3], [4]

1.1.1 Normalizace

Cílem normalizace je sjednotit původní data do stejného měřítka, obvykle do intervalu $[0, 1]$ a zajistit tak stejnou váhu všech dat vstupujících do algoritmu. Oproti standardizaci funguje lépe i na nerovnoměrně rozložená data. Existuje více metod pro normalizaci dat jako například: metody založené na minimálních a maximálních hodnotách, normalizace desetinného škálování či paretovo škálování. Níže jsou stručně popsány nejpoužívanější formy normalizace. [5], [6], [7]

Min-Max škálování

Metoda škáluje data na základě minimálních a maximálních hodnot v rámci daného příznaku. Výpočet normalizovaných dat je dán vztahem:

$$x'_{i,n} = \frac{x_{i,n} - \min(x_n)}{\max(x_n) - \min(x_n)}, \quad (1.1)$$

$x'_{i,n}$ je normalizovaná hodnota i -té hodnoty n -tého příznaku, $x_{i,n}$ je naměřená hodnota, dolní a horní mez z daného příznaku značí proměnné $\min(x_n)$ a $\max(x_n)$.

1.1.2 Standardizace

Úlohou standardizace je sjednotit data do nějakého okolí, pokud vstupní datové jednotky mají vysoké rozdíly mezi jejich rozsahy. Oproti normalizaci je efektivnější pro úpravu dat, které mají gaussovské rozložení. Všeobecně standardizace transformuje střední hodnotu dat na nulu s jednotkovou směrodatnou odchylkou. [6], [7]

Z-skóre

Z-skóre udává poměr střední hodnoty a směrodatné odchylky. Každá nová hodnota $x'_{i,n}$ se z té původní $x_{i,n}$ spočítá jako:

$$x'_{i,n} = \frac{x_{i,n} - \mu_n}{\sigma_n}, \quad (1.2)$$

μ_n je střední hodnota dat v dané skupině a σ_n je směrodatná odchylka, index i značí hodnotu a n daný příznak.

Středně centrovaná standardizace

Známá také jako nulové centrování, má za výsledek posunutí průměru datové sady na nulu, což je popsáno vztahem:

$$x'_{i,n} = x_{i,n} - \mu_n. \quad (1.3)$$

Odečte od každého původního bodu $x_{i,n}$ střední hodnotu μ_n datové sady, výsledkem je $x'_{i,n}$ standardizovaná hodnota i -té hodnoty n -tého příznaku.

Paretovo škálování

Tato normalizace, na základě Paretova pravidla, zavádí myšlenku, že při škálování přikládá větší důležitost datům z větší variabilitou. Využívá se druhá odmocnina směrodatné odchylky podle vztahu:

$$x'_{i,n} = \frac{x_{i,n}}{\sqrt{\sigma_n}}, \quad (1.4)$$

$x'_{i,n}$ je standardizovaná hodnota i -té hodnoty n -tého příznaku, $x_{i,n}$ je původní hodnota a σ_n značí směrodatnou odchylku n -tého příznaku.

1.2 Klasifikační metody

Obecné využití klasifikačních metod v ML již bylo zmíněno. Konkrétněji v medicíně nachází klasifikace uplatnění pro diagnostiku onemocnění, predikci rizika a prognózy na přežití pomocí regresní analýzy, nebo doporučení personalizované léčby pacienta na základě aktuálního stavu. Níže jsou popsány základní klasifikátory učící se pod dohledem, některé z nich budou dále použity v praktické části. [8], [9]

1.2.1 Logistická regrese

Tento klasifikační algoritmus, dále jen LR, využívá sigmoidální funkci, která vybraným příznakům určí pravděpodobnost přidružení do jedné ze dvou skupiny. Na základě této pravděpodobnosti dojde k binární klasifikaci. Pokud je pravděpodobnost rovna nebo vyšší než 0,5 vzorek je klasifikován jako 1 (daná skupina). [10], [11]

1.2.2 Algoritmus K- nejbližších sousedů

Algoritmus nejbližších sousedů, dále jen K- NN, pracuje na základě metrik vzdáleností mezi jednotlivými datovými body. Pro algoritmus se nastavuje parametr k , který vyjadřuje počet nejbližších sousedů. Algoritmus vezme nový datový bod, který chceme zařadit, k tomuto bodu spočítá vzdálenost (nejčastěji se používá euklidovská vzdálenost) na základě příznaků k ostatním bodům z trénovací sady. Podle stanoveného parametru k , přiřadí bod do skupiny k nejbližších sousedů s nejmenší vzdáleností a predikuje tak výstupní třídu tohoto bodu. Nevýhoda může být vysoká citlivost na přeučení modelu a odlehlé hodnoty. [12], [13]

1.2.3 Náhodná doménová struktura

Klasifikátor náhodné doménové struktury, dále jen RF, je založen na náhodných rozhodovacích stromech. Rozhodovací stromy fungují na principu rozhodovacích uzlů, kde se nachází rozhodovací pravidla, například určité prahové hodnoty. Podle odpovědi na uzlovou otázku je zvolena větev ANO/ NE. Poslední uzly, které značí výsledek (predikci třídy) se označují jako listové uzly. V algoritmu RF je uživatelsky nastaveno z kolika náhodných stromů bude predikována výsledná třída na základě většinového pravidla. [14], [15]

1.3 Vyhodnocení správnosti predikce

Tyto metriky vyjadřují výkon modelu předpovězených tříd v porovnání správného zařazení vzorků. Výkon daného modelu lze využít k úpravě parametrů pro lepší klasifikaci. Výběr jednotlivých metrik závisí na konkrétní úloze a některé nejčastěji využívané jsou popsány níže. Pro lepší vyjádření metrik je potřeba si zavést pojmy ohledně pozitivnosti/ negativnosti vzorků. Skutečně pozitivní, dále jen TP , jsou vzorky, které klasifikátor určil jako pozitivní a vzorky jsou skutečně pozitivní. Falešně pozitivní, dále jen FP , jsou vzorky, které model určil jako pozitivní, ale ve skutečnosti spadají pod třídu negativní. Skutečně negativní, dále jen TN , jsou vzorky predikované jako negativní a skutečně spadají do skupiny negativní. Falešně negativní, dále jen FN , jsou vzorky, které klasifikátor určil jako negativní, ale ve skutečnosti patří do skupiny pozitivních. [16]

1.3.1 Přesnost

Přesnost (*accuracy*) udává poměr mezi počtem správně predikovaných vzorků a počtem všech vzorků. Tento poměr se spočítá vztahem:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1.5)$$

Tato metrika není vhodná pro použití na nerovnoměrně rozložené třídy. [17]

Přesnost (*precision*) udává správnost pozitivní třídy. Je to poměr mezi správně určenými pozitivními vzorky a všemi vzorky, které určil klasifikátor jako pozitivní. Tento poměr se spočítá vztahem:

$$precision = \frac{TP}{TP + FP}. \quad (1.6)$$

1.3.2 Senzitivita

Určuje poměr mezi správně klasifikovanými pozitivními vzorky a všemi pozitivními vzorky, který se spočítá vztahem:

$$\textit{senzitivita} = \frac{TP}{TP + FN}. \quad (1.7)$$

Využívá se zejména v případech, kdy chyby ve správnosti mají vážné následky, například v lékařství, kdy klasifikátor zdravého člověka určí jako nemocného. [17]

1.3.3 Specificita

Určuje poměr mezi správně klasifikovanými negativními vzorky a všemi skutečně negativními vzorky. Tento poměr se spočítá vztahem:

$$\textit{specificita} = \frac{TN}{TN + FP}. \quad (1.8)$$

1.3.4 F1 - skóre

Vyjadřuje harmonický průměr mezi přesností (*precision*) a senzitivitou. Čím vyšší výsledná hodnota *f1*, tím je lepší klasifikátor. Harmonický průměr se spočítá vztahem:

$$f1 = 2 \times \frac{\textit{přesnost} \times \textit{senzitivita}}{\textit{přesnost} + \textit{senzitivita}}. \quad (1.9)$$

Tato metrika je vhodná pro použití v případech nevyvážených skupin, aby nedošlo k zaměření modelu jen na jednu skupinu. [18]

2 Běžné metody pro výběr příznaků

Tyto metody jsou založené na postupech, které využívají všechny informace obsažené v datasetu. Níže jsou popsány nejhojněji používané klasické postupy pro extrakci příznaků. [19], [20]

2.1 Filtrační metody

Filtrační metody jsou založeny na výpočtu skóre prostřednictvím statistických závislostí pro jednotlivé příznaky a zachovávají jen ty s nejvyšším skóre na základě prahové hodnoty. [21], [22]

2.1.1 Pearsonův korelační koeficient

Určuje lineární závislost mezi dvěma párovými příznaky v rozmezí $[-1, 1]$. Pokud máme korelaci vysokou (většinou $> 0,7$), znamená to, že oba příznaky přináší podobnou informaci do modelu. V rámci snahy o co nejjednodušší výpočetní náročnost není potřeba oba zachovávat. Vzájemná korelace mezi příznaky se spočítá:

$$r = \frac{[n(\sum xy) - (\sum x \sum y)]}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}, \quad (2.1)$$

n je počet párových pozorování, x a y vyjadřují párové příznaky a $\sum xy$ vyjadřuje sumu součinů těchto párových hodnot. [23], [24]

2.1.2 Chí kvadrát test

Tato statistická metoda se používá k vyjádření vztahu mezi příznakem a jeho odhadovanému výstupu. Čím je hodnota chí kvadrátu větší, tím větší závislost má daný příznak na výstupní proměnné a je vhodné jej zachovat. [25]

2.2 Obálkové metody

Jsou metody založené na algoritmech ML, kterými se snažíme přizpůsobit danému datasetu. Významnost příznaků se určuje na základě lineárních modelů či modelů založených na stromech. Stanovujeme určitou prahovou hodnotu, na základě které vybíráme příznaky. Hodnota prahu se liší podle způsobu aplikace, v regresních úlohách používáme p -hodnotu, zatím co u klasifikačních úloh nám poslouží $f1 - skóre$. [26], [27]

2.2.1 Dopředné prohledávání

Při dopředném prohledávání začínáme tzv. s čistým štítem. Začínáme s prázdnou množinou vybraných příznaků a přidáváme postupně jednotlivé příznaky, které poskytují nejlepší zlepšení výkonu modelu. Dopředné prohledávání probíhá až do stanoveného kritéria ukončení algoritmu např. vybraný počet funkcí, konečný počet iterací či nezlepšování výkonu modelu.

2.2.2 Zpětná eliminace

Zpětná eliminace je založena na postupném odstraňování nejméně významných příznaků. Začínáme s vektorem všech příznaků a postupně je eliminujeme až do stanovených kritérií, která jsou stejná jako u dopředného prohledávání.

2.3 Redukce dimenzionality

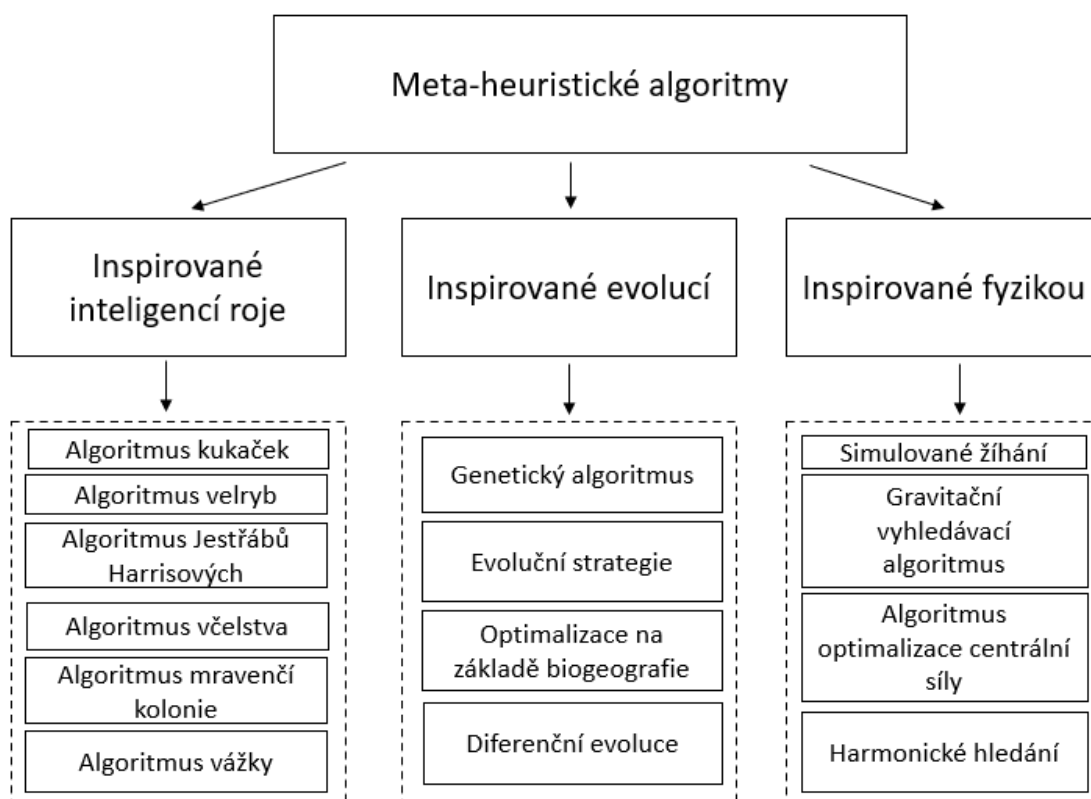
Účelem snížení dimenzionality v datové sadě je zachování co největšího informativního obsahu pro další zpracování, ale snížení výpočetní náročnosti. Důvodem je lepší vizualizace, usnadnění učení modelu a celkového zjednodušení modelu. [28]

2.3.1 PCA

Jedná se o jednu z nejpoužívanějších metod pro redukci dimenzionality, která převádí korelované příznaky na nekorelované příznaky. Základním předpokladem je, že informační váhu dat lze vyjádřit jejich rozptylem. Využívá ortogonální osy nebo-li hlavní komponenty, které určují míru rozptylu daných příznaků. Největší variabilitu nám zachycuje první hlavní komponenta a maximální rozptyl, který je kolmý na variabilu zachycuje komponenta druhá. Prvním krokem na upravených (standardizovaných) datech je výpočet kovariance, která určuje variabilitu mezi příznaky. Dalším krokem je výpočet vlastních čísel a vektorů kovarianční matice. Obecně PCA využívá lineární transformaci ideálně s co největším rozptylem. Finální data dostaneme nafitováním původního datasetu na osy hlavních komponent, kde na každou komponentu mají původní příznaky určitý vliv. [29], [30]

3 Meta-heuristické algoritmy

Základem těchto algoritmů je kompromis mezi procesem randomizace a prohledávání v lokálním okolí daného řešení. Podstatou každého metaheuristického algoritmu jsou diverzifikace a intenzifikace. Pojem diverzifikace objasňuje vytvoření více možných řešení, což umožňuje prohledávat na bázi globálního optima a zabraňuje tak uváznutí řešení pouze v lokálním optimu. Pojem intenzifikace vyjadřuje hledání lokálního okolí pro nalezení lokálního optima. Meta-heuristické algoritmy jsou inspirované inteligencí roje, evolucí a fyzikou, toto rozdělení je naznačené na obrázku 3.1. Níže budou popsány nejčastěji využívané algoritmy, přičemž některé z nich budou dále aplikovány v praktické části této práce. [31], [32], [33]



Obr. 3.1: Rozdělení meta-heuristických algoritmů, převzato z [34]

3.1 Metaheuristické algoritmy inspirované inteligencí roje

Intelligence roje je založená na chování jedinců, kteří se v přírodě vyskytují ve společenstvech, hejnech, koloniích. Jedinci zachovávají v koloniích určitou hierarchii se specifickými úkoly. Algoritmy mají za cíl napodobit jejich chování a řešit tak optimalizační problémy v ML. [35], [36]

3.1.1 Algoritmus podle kukaček

Tento algoritmus byl navržen v roce 2009 pro zpracování obrazů, prognózy a výběr příznaků. Algoritmus podle chování kukaček, dále jen CS, je inspirovaný jak název vypovídá ptačími parazity. V průběhu vývoje algoritmu byl vylepšen o tzv. Levyho let, který odpovídá randomizovanému globálnímu prohledávání. Základní myšlenkou je skutečnost, že každá kukačka snese vejce do náhodně vybraného hnízda. Vejce v hnízdě reprezentuje řešení pro jednoho jedince v populaci, vejce kukačky je analogií nového řešení. Nejlepší hnízda s vysokou kvalitou vajec se přenesou do další generace (nové dobré řešení). Počet možných hostitelských hnízd je pevně daný. Hostitelský pták identifikuje, že se jedná o parazitní vejce s určitou pravděpodobností závisící na podobnosti hostitelských a kukaččích vajec (podobnost optimálního řešení). Při rozpoznání kukaččích vajec může dojít k jejich vyhození či postavení nového hnízda, což znamená nepřípustnost tohoto řešení. Počet hnízd je předem daný a odpovídá velikosti populace. Pro zjednodušení se uvádí, že v každém hnízdě je pouze jedno vejce. Dále je potřeba stanovit délku kroku a směr letu kukačky pomocí Levyho letu. Ten je ekvivalentem náhodné cesty, která umožňuje větší kroky v procesu vyhledávání, jelikož se jedná o distribuci s nekonečným rozptylem. Nové řešení jedince i je dáno rovnicí:

$$X_i(t+1) = X_i(t) + \alpha \otimes L(s, \lambda), \quad (3.1)$$

$\alpha > 0$ odpovídá škálovacímu faktoru dle váhy řešeného problému, X je řešení v čase t a L popisuje Levyho let, který je popsán vztahem:

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin\left(\frac{\pi\lambda}{2}\right)}{\pi} \cdot \frac{1}{s^{1+\lambda}}, \quad (3.2)$$

s odpovídá kroku, Γ je faktorová funkce a λ odpovídá distribuci Levyho letu, nejběžněji $1 < \lambda \leq 3$. [37], [38], [39], [40], [41]

3.1.2 Algoritmus podle jestřábů Harrisových

Algoritmus inspirovaný loveckým chováním jestřábů Harrisových, dále jen HHO. Základem tohoto optimalizačního řešení je tzv. překvapivý skok, při kterém se několik

jedinců jestřábů vrhá koordinovaně na kořist (budeme předpokládat, že kořistí je králík). Jestřábi v tomto algoritmu představují kandidáty na řešení a králík je synonymem optimálního řešení. Jestřábi jsou náhodně rozmístění a čekají na kořist, převedeno do algoritmu, okruh řešení je náhodný a hledá se nejlepší možné řešení. Každý jestřáb, kandidát na řešení, má šanci q na ulovení kořisti, závisující na rozmístění ostatních členů hejna a na vzdálenosti od kořisti. Pravděpodobnost pro každé řešení se mění s každou iterací. Algoritmus se dělí do dvou fází, fáze průzkumu a útočná strategie, přechod mezi nimi souvisí i s polohou králíka a jeho energií na útěk. [42], [43], [44], [45], [46]

Fáze průzkumu

Fáze průzkumu začíná hledáním kořisti v různých oblastech vyhledávacího prostoru a tím pádem i možností objevit potencionální optimální řešení. Při nalezení polohy králíka ostatní jestřábi upravují své pozice podle vůdce, který je králíkovi nejbližší (kandidát na nové optimální řešení). Průzkumná fáze je popsány vztahy:

$$X(t+1) = \begin{cases} X_r(t) - r_1 |X_r(t) - 2r_2 X(t)| & q \geq 0,5 \\ (X_{\text{rabbit}}(t) - X_m(t)) - r_3 (\text{lb} + r_4(\text{ub} - \text{lb})) & q < 0,5 \end{cases}, \quad (3.3)$$

$X(t)$ určuje nynější pozici jestřábů, $X(t+1)$ vyjadřuje pozici jestřábů v další iteraci, $X_{\text{rabbit}}(t)$ pozici kořisti, r_1, r_2, r_3, r_4 jsou náhodná čísla v intervalu $[0, 1]$, která se s každou iterací mění, $X_r(t)$ je náhodně vybraný jedinec z lovicích jestřábů, $X_m(t)$ vyjadřuje průměrnou pozici jestřábů v dané iteraci. Odpovídající průměrná pozice v příslušné iteraci se spočítá rovnicí:

$$X_m(t) = \frac{1}{N} \sum_{i=1}^N X_i(t), \quad (3.4)$$

t je konkrétní iterace, N je celkový počet jestřábů účastnících se lovu a $X_i(t)$ je pozice jednotlivých jedinců.

Před útočnou fází je ještě třeba definovat přechodovou fází závisující na energii králíka, která je dána rovnicí:

$$E = 2E_0 \left(1 - \frac{t}{T}\right), \quad (3.5)$$

E definuje energii (schopnost) králíka uniknout, t je aktuální iterace, T vyjadřuje maximální možný počet iterací, E_0 náleží do intervalu $[-1, 1]$, přičemž se E_0 generuje s každou iterací, až do vyčerpání králíka. Aby došlo k útočné fázi musí platit $|E| < 1$, což vyjadřuje, že poloha králíka byla určena. Pokud je $|E| \geq 1$ pokračuje fáze průzkumu, jelikož králík ještě nebyl spatřen.

Útočná strategie

Tato fáze spočívá v ulovení králíka, jehož poloha je sice určena, ale jako většina kořisti se snaží svým lovcům uniknout. Analogicky optimální řešení je stanoveno, ale může se měnit v závislosti na volbě parametrů či pozici blížících se dobrých řešení. Pravděpodobnost útěku králíka se určuje parametrem r , který je v rozmezí $[0, 1]$. Pokud je $r < 0,5$, lze předpokládat úspěšný útěk králíka a nutnost aktualizace pozic jestřábů na základě jeho pozice Levyho letem. Pokud je $r \geq 0,5$ králík není schopen uniknout, blížíme se k optimálnímu řešení, kterého bude v dalších iteracích dosaženo. Útočnou strategii můžeme rozlišit na dva typy podle síly nebo-li velikosti únikové energie králíka, a to na měkké a tvrdé obléhání.

Měkké obléhání

Měkké obléhání znamená snahu jestřábů unavit králíka, jelikož jeho energie je vysoká $|E| \geq 0,5$, ale tendence k útěku je malá $r \geq 0,5$. Měkké obléhání je dáno rovnicí:

$$X(t+1) = (X_{\text{rabbit}}(t) - X(t)) - E \cdot (\text{skok} \cdot X_{\text{rabbit}}(t) - X(t)), \quad (3.6)$$

s každou další iterací a únikovým skokem králíka klesá parametr $|E|$, dokud nedojde k unavení kořisti (konvergence k 0). Možnou modifikací je měkké obléhání se střídajícími rychlými nálety, při kterých králík využívá četných nečekaných skoků. Tím, že dochází ke střídavým náletům všech jedinců, je zahrnuta jejich průměrná pozice $X(t).mean$. Aktualizace jedince do další iterace $X(t+1)$ je popsána vztahem:

$$X(t+1) = X_{\text{rabbit}}(t) - E \cdot (\text{skok} \cdot X_{\text{rabbit}}(t) - X(t).mean). \quad (3.7)$$

Pokud má králík stále dostatek energie $|E| \geq 0,5$ a vysokou šanci na úspěšný útěk $r < 0,5$, zahajují jestřábi postupné plynulé ponory k unavení kořisti. Ty jsou popsány rovnicí:

$$X(t+1) = X_{\text{rabbit}}(t) - E \cdot (\text{skok} \cdot X_{\text{rabbit}}(t) - X(t)). \quad (3.8)$$

Při neúspěšném lovu dochází k aktualizaci polohy jedinců náhodným krokem Levyho letu.

Tvrdé obléhání

V případě kdy šance na úspěšný únik je malá $r \geq 0,5$ a králík je oslaben $|E| < 0,5$ zahajují jestřábi nálety střežhlav dolů pro co nejrychlejší chycení kořisti. Tvrdé obléhání je popsáno vztahem:

$$X(t+1) = X_{\text{rabbit}}(t) - E \cdot (X_{\text{rabbit}}(t) - X(t)). \quad (3.9)$$

3.1.3 Algoritmus podle velryb

Jedná se o poměrně nový algoritmus, představený v roce 2016, inspirovaný chováním velryb při hledání potravy v roji, dále jen WOA. Za zmínku stojí, že WOA funguje na podobném principu jako další z algoritmů inspirovaných rojem a to algoritmus vlčí smečky. WOA je silný ve vyhledávání v globálním optimu, avšak může velmi snadno dojít k uvíznutí na lokálním optimu. Je založen na způsobu lovení ve vícerozměrném prostoru. Je potřeba rozmístit jedince po spirále, což odpovídá rozhodovacím proměnným závislejícím na čase. Aktuální poloha jedince je závislá na třech procesech v průběhu algoritmu. Těmi jsou fáze průzkumu, nebo-li hledání kořisti, dále zmenšující se seskupení kořisti a fáze útoku za využití bublinové sítě (fáze těžby). Při algoritmu WOA se předpokládá poloha kořisti jako nejlepší kandidát na řešení, který se blíží ke globálnímu optimu. Snahou je najít velrybu, která má ke kořisti nejbližší a ostatní podle ní aktualizují své pozice. Poloha každého jedince je popsána vektory, které popisují dimenzi prostoru D . Jak již bylo zmíněno kepokraci se pohybují po spirále aby co nejlépe obklíčili kořist. [47], [48]

Fáze průzkumu

Pátrací velryba hledá kořist (nejlepší řešení) na základě polohy ostatních jedinců. Umístění velryb v aktuální iteraci t je popsáno rovnicemi:

$$D = |C \cdot X_r - X|, \quad (3.10)$$

$$X(t + 1) = X_r - A \cdot D, \quad (3.11)$$

X_r je náhodná poloha pátrací velryby, X je pozice velryby, vůči němuž se pohybují ostatní pátrací jedinci. Parametry A a C řídí jejich pohyb v prohledávaném prostoru a jsou zadány vztahy:

$$A = 2a \cdot r - a, \quad (3.12)$$

$$C = 2r, \quad (3.13)$$

r je náhodná proměnná v rozsahu $[0, 1]$ a a je lineárně klesající s počtem iterací od 2 k 0.

Fáze obklopování kořisti

Velryby postupně aktualizují své polohy a obklopují kořist podle vztahů:

$$D = |C \cdot X^*(t) - X(t)|, \quad (3.14)$$

$$X(t+1) = X^*(t) - A \cdot D. \quad (3.15)$$

Za nejlepší řešení se považuje jedinec který je blízko kořisti X^* v aktuální iteraci t , k němuž aktualizují své polohy ostatní X .

Fáze útoku

Útok na kořist probíhá pomocí bublinkové sítě ve spirále, který donutí se kořist shluknout do blízkého okolí. Pohyb po spirále je určen rovnicí:

$$X(t+1) = D'' \cdot e^{bl} \cdot \cos(2\pi l) + X^*, \quad (3.16)$$

D'' je vzdálenost mezi nejbližší velrybou a kořistí, l je náhodné číslo spadající do intervalu $[-1, 1]$, konstanta beta b definuje tvar logaritmické spirály a X^* je dosud nejlepší řešení. Tento pohyb společně s aktualizací pozic velryb vůči nejlepšímu řešení definuje styl jejich plavání na základě pravděpodobnosti p . Pravděpodobnost určí zda dojde k pohybu po spirále, jelikož je v dohledném okolí nejlepší řešení či bude nadále prováděna fáze průzkumu. Pohyb velryb je popsán rovnicemi:

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & p < 0.5 \\ D'' \cdot e^{bl} \cdot \cos(2\pi l) + X^* & p \geq 0.5 \end{cases}. \quad (3.17)$$

3.1.4 Optimalizace podle papouška

Nový algoritmus navržený v roce 2024, který napodobuje společenské chování oblíbeného domestikovaného papouška *Pyrura zelenolícího*, dále jen PO. Algoritmus je založen na napodobení čtyř stavů, kterými jsou krmení, udržení se na chovateli, komunikace mezi jedinci a úlet při strachu z nepřítele. [49]

První stav potravního chování na základě odhadu umístění krmiva, které je zadáno vztahem:

$$X_i(t+1) = (X_i(t) - X_{best}) \cdot L(s, \lambda) + r \cdot \left(1 - \frac{t}{Max_{iter}}\right)^{\left(\frac{2t}{Max_{iter}}\right)} \cdot X_{mean}(t). \quad (3.18)$$

Pozice v další iteraci pro i -tého jedince $X_i(t+1)$, X_{best} prozatím neoptimálnější pozice papouška (analogie nejlepšího řešení), $L(s, \lambda)$ je Levyho let v prohledávaném prostoru řešení, r je náhodné číslo z intervalu $[0, 1]$ a $X_{mean}(t)$ značí průměrnou

pozici hejna s počtem jedinců N v aktuální iteraci t . Průměrná pozice se spočítá jako:

$$X_{mean}(t) = \frac{1}{N} \sum_{i=1}^N X_i(t). \quad (3.19)$$

Dalším typickým chováním je sezení na chovateli, které je popsáno rovnicí:

$$X_i(t+1) = X_i(t) + X_{best} \cdot L(s, \lambda) + r \cdot o, \quad (3.20)$$

$X_i(t+1)$ vyjadřuje pozici pro i -tého jedince v následující iteraci, X_{best} prozatím nejlepší pozice papouška, $L(s, \lambda)$ je Levyho let, který vyjadřuje let k chovateli. Parametry, které společně vyjadřující postavení papouška na různé části těla chovatele jsou r , což je náhodné číslo z intervalu $[0, 1]$ a o vyjadřující chvilkovou nehybnost papouška na chovateli vektorem jedniček závislejícím na velikosti dimenze prohledávaného prostoru. Jelikož je *Pyrura* velmi společenský papoušek, nastává i stav komunikace v hejnu. Chování zahrnuje dva stejně pravděpodobné stavy a těmi jsou přilet papouška za účelem komunikace s ostatními jedinci při pravděpodobnosti $p \leq 0,5$ a odlet jedince z hejna $p > 0,5$. Aktualizovaná pozice X_i pro i -tého jedince v aktuální iteraci t na základě pravděpodobnosti je popsána rovnicemi:

$$X_i(t+1) = \begin{cases} 0,2 \cdot r \cdot \left(1 - \frac{t}{Max_{iter}}\right) \cdot (X_i(t) - X_{mean}(t)) & p \leq 0,5 \\ 0,2 \cdot r \cdot \exp\left(-\frac{t}{r \cdot Max_{iter}}\right) & p > 0,5 \end{cases}, \quad (3.21)$$

při letu do hejna uvažujeme průměrnou pozice hejna X_{mean} . Posledním možným chováním je úlet při strachu z nepřítele, které je vyjádřeno vztahem:

$$X_i(t+1) = X_i(t) + r \cdot \cos\left(0,5\pi \cdot \left(\frac{2t}{Max_{iter}}\right)\right) \cdot (X_{best} - X_i(t)). \quad (3.22)$$

3.1.5 Algoritmus opylování květin

Jak už název vypovídá, optimalizace napodobuje proces opylování květin, dále jen FPA. V rámci opylení mezi rostlinami uvažujeme při přenesení pylových zrn Levyho let či přenosem za pomocí živočichů (hmyz). Platí zde také zjednodušení, že každá květina má pouze jeden květ a jednu pylovou gametu (analogie možného řešení), uvažujeme pouze jeden druh rostlin. Podle pravděpodobnosti p určujeme zda dochází ke globálnímu či lokálnímu prohledávání. Pokud je náhodně zvolené číslo $r < p$, pak se jedná o globální opylení se zavedenou náhodností Levyho letu $L(s, \lambda)$. Globální opylení je zadáno jako:

$$X_i(t+1) = X_i(t) + L(s, \lambda) \cdot (X_{best} - X_i(t)), \quad (3.23)$$

aktualizovaná pozice opylené gamety (řešení) X_i je dána pozicí i -tého jedince v aktuální iteraci t s přihlédnutím na dosud nejlepší řešení X_{best} .

Pokud je náhodně zvolené číslo $r \geq p$, jedná se o lokální opylení zadané vztahem:

$$X_i(t+1) = X_i(t) + \varepsilon \cdot (X_{r_1} - X_{r_2}). \quad (3.24)$$

Berou se v úvahu dva náhodní jedinci z populace X_{r_1} a X_{r_2} , jejichž možnost křížení je ovlivněna parametrem uniformní distribuce ε z rozsahu $[0, 1]$. [50], [51]

3.1.6 Algoritmus umělého včelstva

Tento algoritmus navržený v roce 2005 je inspirovaný chováním včel medonosných v koloniích, dále jen ABC. V tomto algoritmu existují tři druhy včel: dělnice, průzkumné a přihlížející včely. Dělníci vybírají potravu a sdílí informace s přihlížejícími včelami, které vybírají pouze dobré zdroje potravy. Kvalitnější zdroj má větší šanci na vybrání (nové dobré řešení). Průzkumné včely se vymaní ze skupiny dělníků a letí hledat novou (vhodnější) potravu. Počet zdrojů potravy se rovná počtu dělníků. ABC můžeme rozdělit do čtyř fází a to iniciační fáze, dělnická fáze, fáze přihlížející včely a fáze průzkumné včely. Inicializační fáze pro každý zdroj potravy nebo-li možné řešení X je zadána:

$$X_{i,j} = X_j^{min} + r \cdot (X_j^{max} - X_j^{min}), \quad (3.25)$$

i je náhodně vygenerovaná velikost zdrojů potravy a j reprezentuje dimenzi. Parametr r je náhodné číslo z intervalu $[0, 1]$. Po inicializační fázi vytvoření zdrojů potravy přichází na řadu dělnická fáze, která je popsána rovnicí:

$$V_{i,j} = X_{i,j} + \phi_{i,j} \cdot (X_{i,j} - X_{k,j}). \quad (3.26)$$

Nový zdroj potravy označujeme jako V v dané dimenzi podle j , $\phi_{i,j}$ je náhodné číslo z intervalu $[-1, 1]$ a k značí náhodně vybraný zdroj potravy. Přihlížející včely následně pokračují v hledání nových zdrojů, avšak jen v okolí těch zdrojů potravy, které byly na základě sdílených informací od dělnických včel zvoleny jako vhodné. Pravděpodobnost, že nový zdroj bude vybrán jako kvalitní, závisí na jeho hodnotě (množství nektaru) a na počtu zdrojů potravy SN ekvivalentnímu populaci dělnických včel. Pravděpodobnost vybrání je zadána:

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j}, \quad (3.27)$$

p_i je pravděpodobnost výběru potravy s odpovídajícím množstvím nektaru fit . Pro kvalitu jednotlivých zdrojů platí:

$$fit_i = \begin{cases} \frac{1}{1+f(X_i)} & \text{když } f(X_i) \geq 0 \\ 1 + |f(X_i)| & \text{jinak} \end{cases}, \quad (3.28)$$

$f(X_i)$ vyjadřuje hodnotu účelové funkce a platí, že pokud nový zdroj je lepší než předešlý, je starý zdroj zapomenut a dále se s ním nepočítá. Poslední fází je fáze včely průzkumné, která má za úkol zkontrolovat kvalitu každého zdroje. Pokud zdroj přesáhne nastavený limit, považujeme ho za vyčerpaný a daný dělník, ke kterému příslušel tento zdroj potravy se změní ve včelu průzkumnou a generuje nový zdroj potravy. [52], [53], [54]

3.1.7 Algoritmus mravenčí kolonie

Tento algoritmus byl inspirován chováním v kolonii mravenců, dále jen ACO. Sociální chování mravenců je založeno na souboru dynamických mechanismů a sebeorganizaci, která vede k udržení globálně optimálního systému. Funkčnost ovlivňují čtyři složky sebeorganizace: náhodnost, vácenásobná obnova, pozitivní a negativní zpětná vazba. Mezi jedinci dochází ke dvěma způsobům komunikace. Přímá komunikace, kde dochází k blízkému kontaktu a výměně potravy. Při vzdálené komunikaci mravenci používají feromony (analogie k datovým informacím), zanechávají je jako značení k nalezení nejkratší cesty mezi mraveništem a potravou. Jelikož se feromony rychle odpařují, dávají mravenci přednost silnějšímu zastoupení feromonů a dochází tak k dynamické modifikaci jejich chování v každé iteraci s předanými informacemi od minulých generací. Chování mravenců popisují dvě pravidla, pravidlo přesunu a pravidlo pro aktualizaci feromonové stopy. Meziřešení daného problému popisuje přesun mravence ze stavu r do stavu s s určitou pravděpodobností pro každého k -tého mravence, která je popsána vztahem:

$$s = \arg_{u \in J_r^k} \max [\tau(r, u)^\alpha \cdot \eta(r, u)^\beta] \quad q \leq q_0. \quad (3.29)$$

Pravděpodobnost přesunu mezi stavy s a r je zadáno rovnicemi:

$$p_k(r, s) = \begin{cases} \frac{\tau(r, s)^\alpha \cdot \eta(r, s)^\beta}{\sum_{u \in J_r^k} \tau(r, u)^\alpha \cdot \eta(r, u)^\beta} & s \in J_r^k \\ 0 & s \notin J_r^k \end{cases}. \quad (3.30)$$

Dojde k posunu v případě nalezení lepšího řešení, pokud je nalezené řešení horší, jeho mezipaměť se resetuje na 0 (vymizení feromonové stopy). Hladinu feromonů mezi stavem r a stavem u v dané generaci vyjadřuje proměnná τ . Délka přesunu mezi těmito stavy je značena η . Nová nenalezená řešení jsou vyjádřena parametrem J_r^k pro každého k -tého mravence v generaci. Jako kontrolní parametry síly feromonů se využívají α a β , jejich odpařování je v rozmezí $[0, 1]$ vyjádřeno parametrem q . Dalším krokem je aktualizace feromonových stop v každé iteraci po vykonání meziřešení všech jedinců v kolonii, tento krok se nazývá pravidlo pro aktualizaci feromonové stopy. Pro optimální řešení je třeba aktualizovat stopy na lokální bázi

definované jako:

$$\tau(r, u) = (1 - \rho)\tau(r, s) + \sum_{k=1}^m \Delta\tau_k(r, s). \quad (3.31)$$

Každý mravenec mezi jednotlivými iteracemi přidává feromony při přechodů z jednoho stavu do druhého. Množství feromonů je popsáno rovnicemi:

$$\Delta\tau_k(r, s) = \begin{cases} \frac{Q}{L_k} & (r, s) \in \pi_k \\ 0 & (r, s) \notin \pi_k \end{cases}. \quad (3.32)$$

Dráha L_k na které mravenec zanechá feromony je vzdálenost posloupnosti π_k za čas $\Delta\tau$, Q je konstantní. [55], [56], [57], [58], [59]

3.1.8 Algoritmus světlušky

Tento algoritmus, dále jen FA, navržený v roce 2008 je inspirován svítivostí světlušky. Na základě schopnosti bioluminiscence přitahují světlušky ostatní jedince či kořist. Oproti reálné přírodě se zavádí zjednodušující předpoklad, že blikání nezávisí na pohlaví. Jasnější světluška, která je analogií k lepšímu řešení, vždy přitahuje méně jasnější jedince, kteří odpovídají neoptimálnímu řešení daného problému. Světluška, která přitáhne nejvíce jedinců se považuje za optimální řešení. Svítivost světlušky pro okolní jedince je zadána vztahem:

$$I(r) = I_0 e^{-\gamma r^2}, \quad (3.33)$$

$I(r)$ je zeslabená původní svítivost I_0 , kterou vidí jedinec ve vzdálenosti r , parametr γ udává absorpci světla. Vzdálenost r mezi dvěma jedinci se počítá na základě euklidovské vzdálenosti:

$$r_{ij} = \|X_i - X_j\| = \sqrt{\sum_{k=1}^d (X_{i,k} - X_{j,k})^2}, \quad (3.34)$$

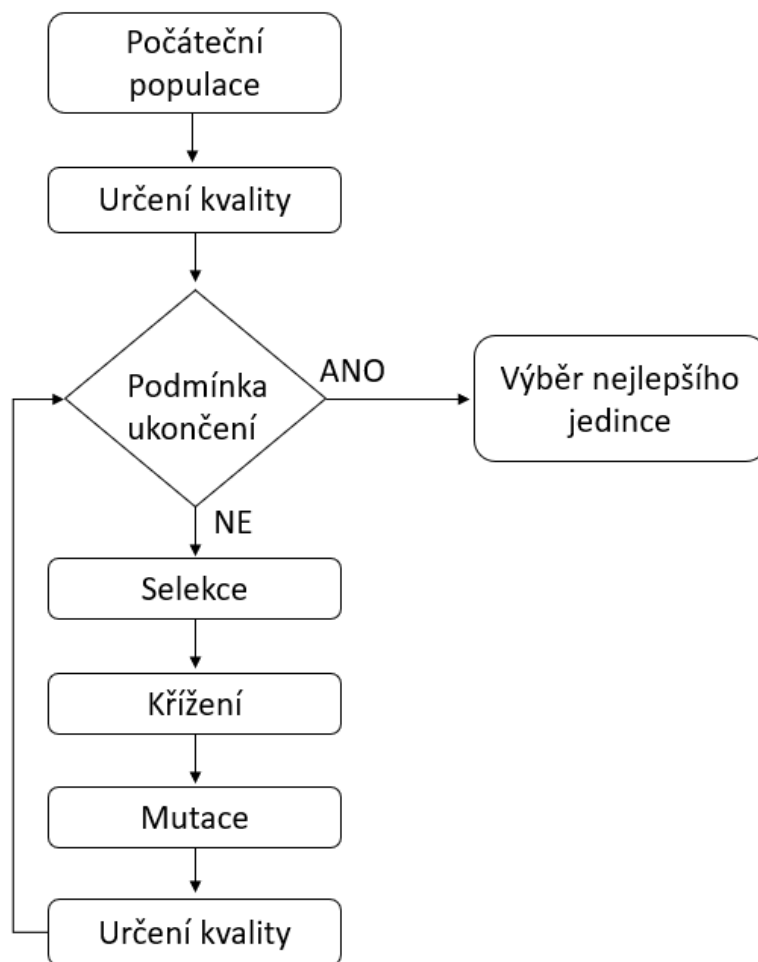
X vyjadřuje možné řešení, indexy i a j značí jedince v dimenzi d . Přitahování méně jasnějšího jedince i více jasnějším jedincem j je popsáno rovnicí:

$$X_i = X_i + \beta_0 e^{-\gamma r_{ij}^2} (X_j - X_i) + \alpha \epsilon_i, \quad (3.35)$$

parametr β_0 odpovídá konstantní svítivosti jedince, α je váha náhodného pohybu ϵ zadaného vektorem náhodných čísel sestaveného z Gausovské či uniformní distribuce. [60], [61]

3.2 Metaheuristické algoritmy inspirované evolucí

Tato skupina algoritmů funguje na principu přirozeného výběru a přežití nejlepších jedinců. Simuluje procesy evoluce jakými jsou genetická variabilita, selekce, křížení a mutace. Jednotlivé kroky obecného evolučního algoritmu naznačuje vývojový diagram 3.2, avšak tato posloupnost kroků nemusí být dodržena u všech algoritmů. Těmito procesy se vytváří nová generace, která je analogicky považována za nové lepší řešení. [62]



Obr. 3.2: Vývojový diagram genetického algoritmu, převzato z [63]

3.2.1 Genetický algoritmus

Tento algoritmus, dále jen GA, napodobuje Darwinovu teorii výběru v populaci. Populace je definována souborem jedinců, kteří představují kandidáty na řešení daného problému. Charakteristika každého jedince je zakódována v genech. Selektce probíhá

na základě kvality jedince, která je rovna kritériální funkci. V rámci selekce máme několik přístupů jako je souboj nebo vážená ruleta. Vybraní jedinci podléhají křížení za vzniku nových jedinců. Noví jedinci podléhají křížení na základě určité pravděpodobnosti. Ukončujícím kritériem může být maximální počet iterací či dosažená kvalita jedince. [64]

Selekce

Souboj

Do souboje vstupují jedinci s jejich kvalitami, náhodně vybírám dva jedince, jejichž kvality porovnám. Do následující generace postupuje kvalitnější jedinec. Provádíme tolik soubojů, kolik máme jedinců, abychom zachovali stejnou velikost nové generace.

Vážená ruleta

Tento výběr je založen na pravděpodobnostním přístupu. Jedinci s vyšší kvalitou mají vyšší pravděpodobnost na postup do další generace. Prvním krokem je určení pravděpodobnosti výběru jedince, kterou vypočteme vydělením kvality jedince součtem všech kvalit. Pro každého jedince se spočítá distribuční funkce, což znamená částečnou sumu (interval, do kterého jedince spadá). Opakujícími dvěma kroky podle počtu jedinců je vygenerování náhodného čísla z intervalu $[0, 1]$ a vybrání jedince do další generace, který svojí distribuční funkcí odpovídá náhodnému číslu.

Křížení

Jednobodové křížení

Ze začátku je zadefinovaná pravděpodobnost křížení. Sekvenčně je procházena matice jedinců a mezi dvěma sousedními jedinci proběhne křížení na základě náhodně vygenerované pravděpodobnosti p z intervalu $[0, 1]$. Pokud je p menší než pravděpodobnost křížení dojde k výměně informace mezi původními jedinci v náhodně zvoleném bodě za vzniku nových dvou potomků, kteří postupují do další generace. Pokud je p vyšší než pravděpodobnost křížení, ke křížení nedojde a do nové generace postupují původní jedinci.

Dvoubodové křížení

Tento postup je podobný jednobodovému křížení avšak změna je ve výměně informace, která neprobíhá jen v jednom náhodném bodě, ale ve dvou náhodně zvolených bodech.

Mutace

Na začátku definujeme hodnotu pravděpodobnosti mutace a generujeme matici pravděpodobnosti mutace v intervalu $[0, 1]$, která má stejný rozměr jako daná populace jedinců. Postupně procházíme matici jedinců a porovnáváme hodnotu s maticí mutace. Pokud je hodnota v matici mutace na dané pozici menší, než pravděpodobnost mutace, dojde k inverzi bitu u jedince na dané pozici.

3.2.2 Diferenciální evoluce

Diferenciální evoluce, dále jen DE, je vhodným evolučním algoritmem pro hledání globálních extrémů. Počáteční populace P je definována jedinci, kteří jsou zakódováni ve vektorech v prohledávacím prostoru. Pro každého aktivního jedince je vygenerován konkurenční (zkušební) jedinec. Do nové generace se zařadí bod s nižší funkcí hodnotou. Konkurenční vektor vzniká z aktivního a šumového vektoru, které jsou podrobeny mutaci z několika bodů staré populace P . [65], [66], [67]

Mutace

Mutace vytváří šumový vektor, který vzniká z původních jedinců. Postupně ke každému jedinci z populace, aktivní jedinec, vybíráme další 3 až 4 jedince, kteří se od sebe odečítají a následně se násobí mutační konstantou v intervalu $[0, 1]$.

Křížení

Po mutaci dochází ke zkřížení aktivního jedince a šumového vektoru za vzniku zkušebního jedince. Na náhodně vybrané počáteční pozici mutace je zkušební jedinec shodný s aktivním jedincem. Následující pozice do vektoru zkušebního jedince jsou vybírány podle hodnoty konstanty křížení. Pokud je náhodně vygenerovaná hodnota menší nebo rovna než hodnota křížení, na odpovídající pozici zkušebního jedince se propisuje hodnota šumového vektoru. Pokud je náhodná hodnota křížení větší než konstanta, do zkušebního jedince vstupuje na odpovídající pozici číslo z aktivního jedince (rodiče).

Selekce

Do souboje vstupují rodiče a jejich zkušební jedinci s jejich kvalitami. Funkční hodnoty těchto vektorů porovnáme a do další generace postupuje kvalitnější jedinec. Provádíme tolik soubojů, kolik máme rodičů, abychom zachovali stejnou velikost nové generace.

3.3 Metaheuristické algoritmy inspirované fyzikou

Tyto algoritmy jsou založeny na jevech, které jsou vysvětlovány pomocí fyzikálních zákonů. [68]

3.3.1 Simulované žíhání

Tento algoritmus je založen na žíhání krystalického tělesa, dále jen SA. Těleso je zahříváno na vysokou teplotu a postupně je ochlazováno s předpokladem snižování defektů v krystalické mřížce. Na základě pravděpodobnosti, která je úměrná teplotě jsou přijímána i horší řešení. Čím více dochází k ochlazení tělesa, tím menší je pravděpodobnost na přijetí horšího řešení. Pravděpodobnost přijetí řešení je dána Metropolisovým kritériem, které je vyjádřeno rovnicemi:

$$P(x' \rightarrow x) = \begin{cases} 1 & f(x') \leq f(x) \\ \exp\left(\frac{-|f(x')-f(x)|}{T}\right) & f(x') > f(x) \end{cases}, \quad (3.36)$$

P je pravděpodobnost nahrazení bodu x za nový bod x' . Výrazy $f(x')$ a $f(x)$ vyjadřují funkční hodnotu v bodech x' a x . Teplota tělesa T klesá s počtem iterací. [69]

4 Datasety a úprava dat

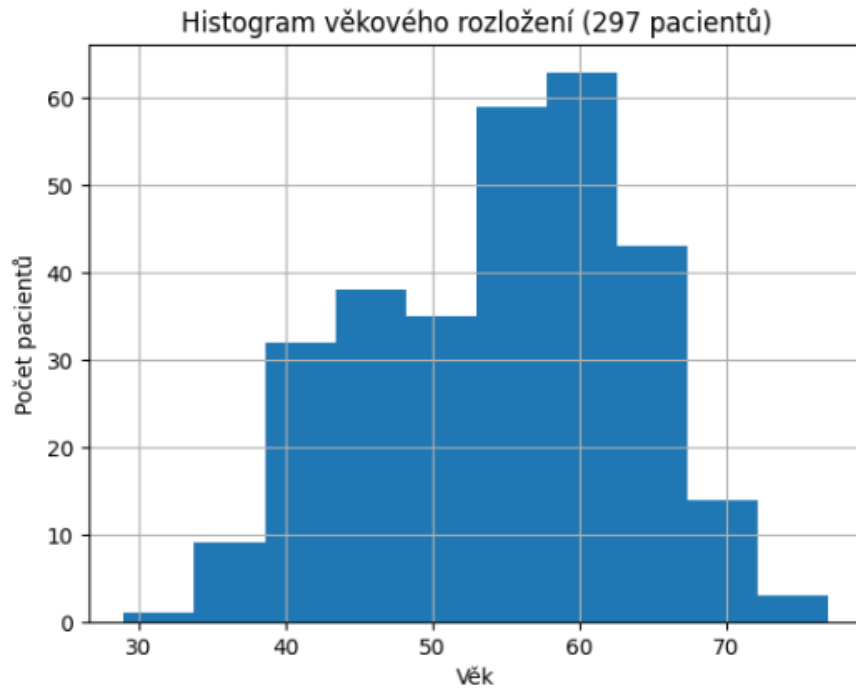
Načtení datasetů, úprava dat, implementace algoritmů a vyhodnocení bylo provedeno v programovém prostředí Python (v3.11.8, [70]).

4.1 Kleveladský dataset srdečního onemocnění

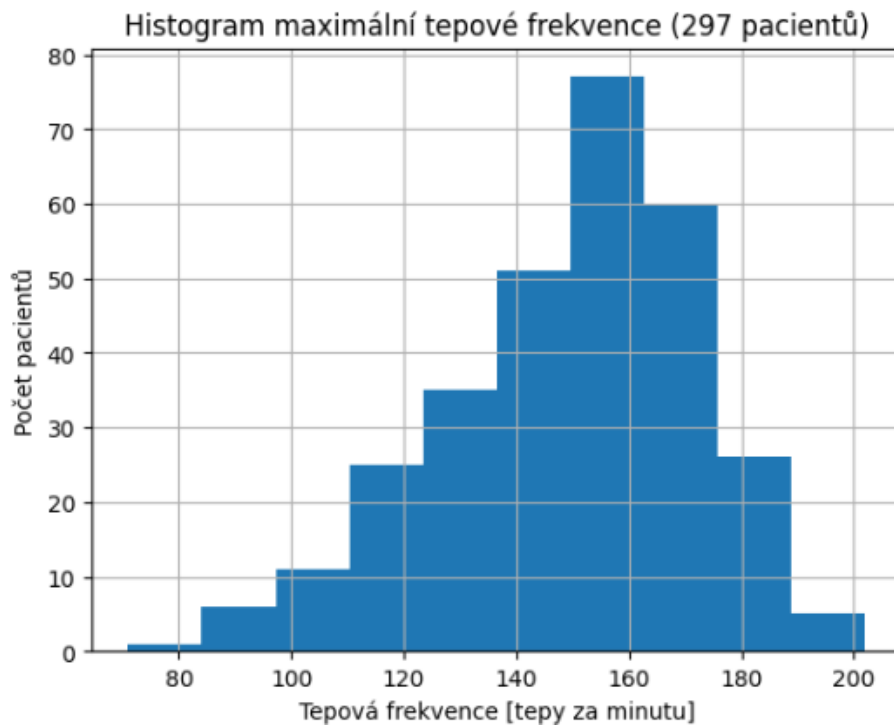
Jako první z datasetů na otestování výběru příznaků za pomoci meta-heuristických algoritmů byl využit Kleveladský dataset z Kalifornské univerzity [71]. Pro tuto databázi existuje v pythonu balíček *ucimlrepo* [72]. Dataset zahrnuje 303 respondentů (206 mužů a 97 žen). Z celkového počtu 76 atributů se pro výzkumy s využitím ML využívá pouze 14 z nich, přičemž je zahrnuta diagnóza zda se onemocnění vyskytuje či nikoliv. Z použitých atributů jsou zde často používané kategorie v klasifikaci onemocnění jako je věk či pohlaví. Ze specifitějších příznaků pro srdečních onemocnění zde vystupuje typ bolesti na hrudi, maximální tepová frekvence nebo také změna ST úseku při zařazení zátěže. Jednotlivé příznaky jsou vypsány v příloze A.1.

4.1.1 Rozložení a úprava dat

Prvním krokem bylo nalezení a náhrada chybějících hodnot, označených jako *NAN* hodnoty. V případě příznaku *ca* chyběly 4 hodnoty a u kategorie *thal* pouze dvě. Při této skutečnosti bylo možné tyto pacienty vypustit z datasetu, jelikož pro klasifikační účely zůstalo stále 297 respondentů ve věku od 29 do 77 let (průměr 54 let). Dalším krokem bylo ověření normálního rozložení dat. Jak již bylo zmíněno v teoretické části, normalita rozložení dat může ovlivnit robustnost modelu a dle toho musíme správně volit další postupy. Klasickým postupem bylo vizuální hodnocení podle histogramů. Na spojité příznaky byl aplikován statistický Shapiro-Wilkův test s hladinou významnosti 0,05, který ve své nulové hypotéze předpokládá, že data pocházejí z normálního rozložení. Příklady některých rozložení jsou na grafech 4.1, 4.2. Nulová hypotéza na základě *p* hodnoty byla zamítnuta, tudíž v dalších krocích bude na dataset použita normalizace, která nepředpokládá normální rozložení. Při analýze odlehlých hodnot se nejevila žádná jako extrémně abnormální. V této úloze nám jde o výběr příznaků z nichž dochází k predikci pouze zdravý/ nemocný, proto bylo dalším krokem skupiny nemocných podle závažnosti spojit pouze do jedné kategorie označené jako 1, tzn. že se u nich onemocnění vyskytuje. Jako další byl pro specifické příznaky, kde se nachází více kategorií, konkrétně *cp*, *slope* a *thal* využit *one – hot encoding*, což znamená transformaci na binární proměnné pro lepší funkčnost algoritmů. Tím pádem z původních 13 příznaků vznikne 20. Posledním krokem z předpracování dat byla normalizace hodnot v příznacích.



Obr. 4.1: Histogram věkového rozložení



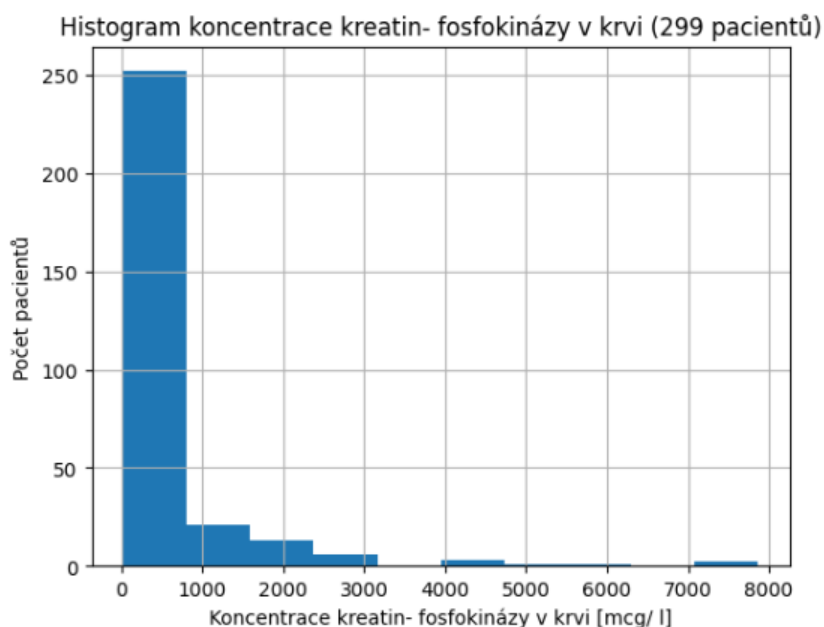
Obr. 4.2: Histogram maximální tepové frekvence [tepy za minutu]

4.2 Dataset srdečního selhání

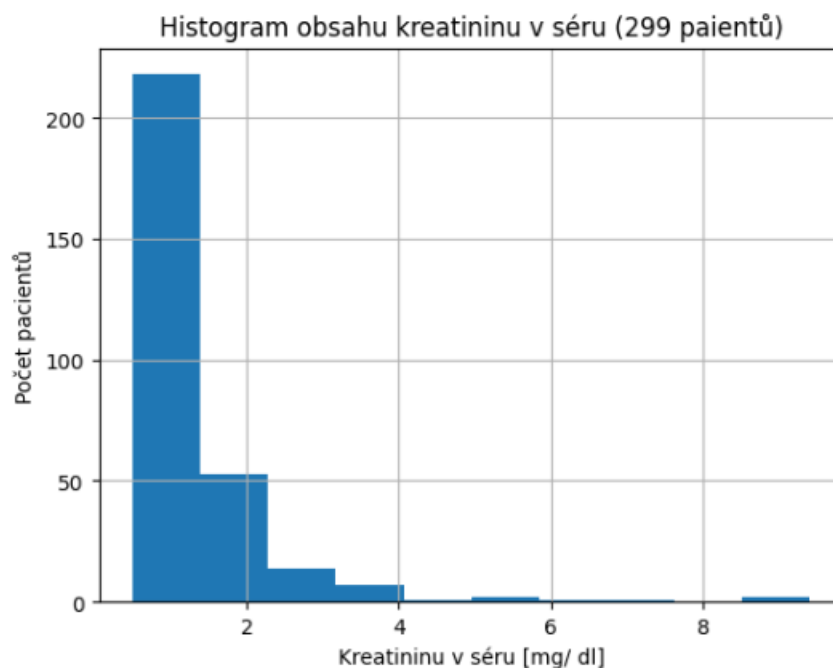
Jako další dataset na otestování výběru příznaků za pomoci meta-heuristických algoritmů byl využit dataset srdečního selhání shromážděného na Faisalabad institutu kardiologie. Zveřejněn je v repozitáři Kalifornské univerzity [73] a existuje pro něj v pythonu balíček *ucimlrepo* [72]. Dataset zahrnuje 299 respondentů ve věkové hranici 40 - 95 let (průměr 61 let) se zastoupením 105 mužů a 194 žen. Pacienti jsou popsáni 13 příznaky, přičemž je zahrnuta diagnóza zda pacient během sledovacího období umřel. Z použitých atributů se zde vyskytují kategorie jako je věk, pohlaví, přítomnost diabetu, hodnota sodíku v krevním séru nebo také zda je pacient kuřák. Jednotlivé příznaky jsou vypsány v příloze B.1.

4.2.1 Rozložení a úprava dat

V datasetu se nenachází žádné chybějící hodnoty. Následovalo ověření normálního rozložení dat, kde byl opět aplikován statistický Shapiro-Wilkův test s hladinou významnosti 0,05. Příklady některých rozložení jsou na grafech 4.3, 4.4. Na základě zamítnutí nulové hypotézy byla provedena normalizace, před níž však došlo k odstranění odlehlých hodnot. Co se týče příznaku kreatin-fosfokinázy, byli odstraněni pacienti s hodnotou přes 3000 [mcg/dl], obsah krevních destiček v krvi byl redukován pod hranici 600000 [počet/ μ l] a obsah kreatininu v séru byl redukován pod 4 [mg/dl]. Po těchto úpravách pro další predikci, zda pacient během sledovacího období zemřel, vstupuje 283 respondentů.



Obr. 4.3: Histogram hodnoty kreatin-fosfokinázy v krvi [mcg/l]



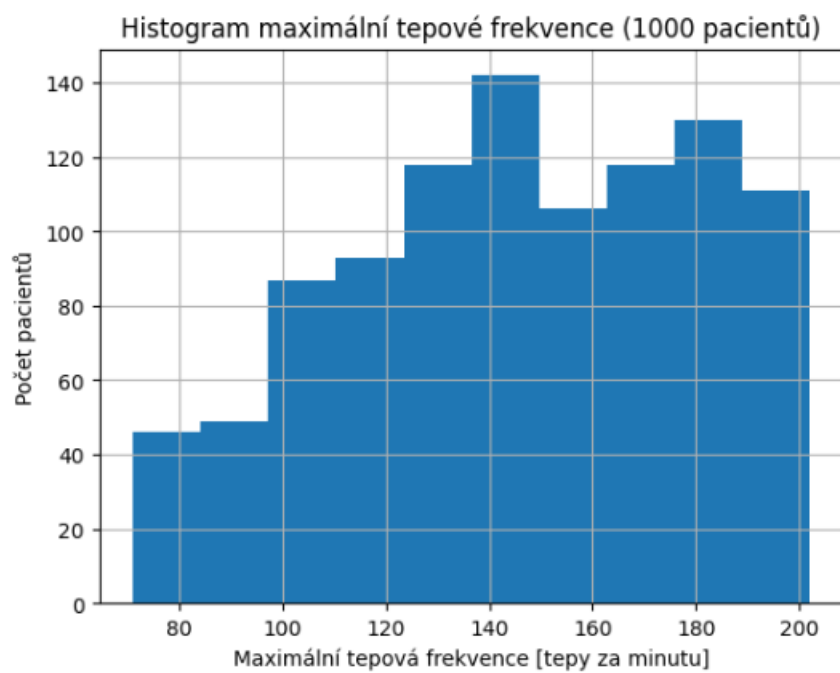
Obr. 4.4: Histogram kreatininu v séru [mg/dl]

4.3 Indický dataset srdečního onemocnění

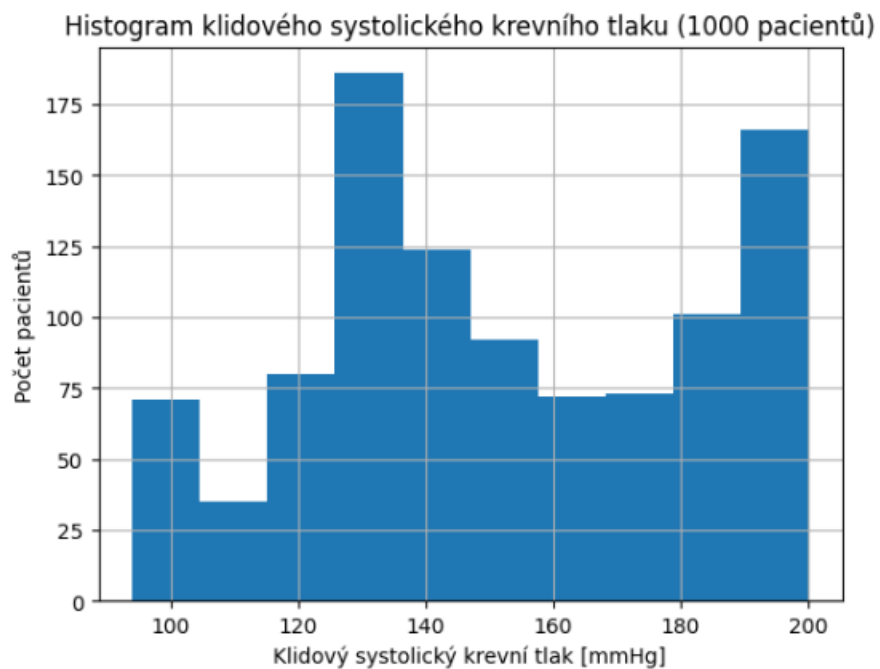
Jako poslední byl pro praktické otestování využit dataset srdečního onemocnění shromážděného v nemocnicích v Indii [74]. Dataset zahrnuje 1000 respondentů ve věkové hranici 20 - 80 let (průměr 49 let) se zastoupením 765 mužů a 235 žen. Ke každému pacientovi je přiřazeno 14 příznaků, přičemž je zahrnuta diagnóza zda pacient má či nemá srdeční onemocnění. První atribut ID pacienta byl vyřazen ještě před další analýzou dat jako irelevantní. Z použitých atributů se zde vyskytují kategorie jako je věk, pohlaví, hodnota cholesterolu v krevním séru nebo také typ bolesti na hrudi. Jednotlivé příznaky jsou vypsány v příloze C.1.

4.3.1 Rozložení a úprava dat

V tomto datasetu se nevyskytují žádné chybějící hodnoty. Na číselné příznaky byl aplikován statistický Shapiro-Wilkův test s hladinou významnosti 0,05 pro ověření normality rozložení dat. Příklady některých rozložení jsou na grafech 4.5, 4.6. Dalším krokem bylo pro specifické příznaky, kde se nachází více kategorií, konkrétně *chestpain*, *slope* a *restingrelectro* využít *one – hot encoding*, což znamená transformaci příznaků na binární proměnné pro lepší funkčnost algoritmů. Tím pádem z původních 12 příznaků (po odstranění ID) vznikne 20. Posledním krokem z předzpracování dat byla normalizace hodnot.



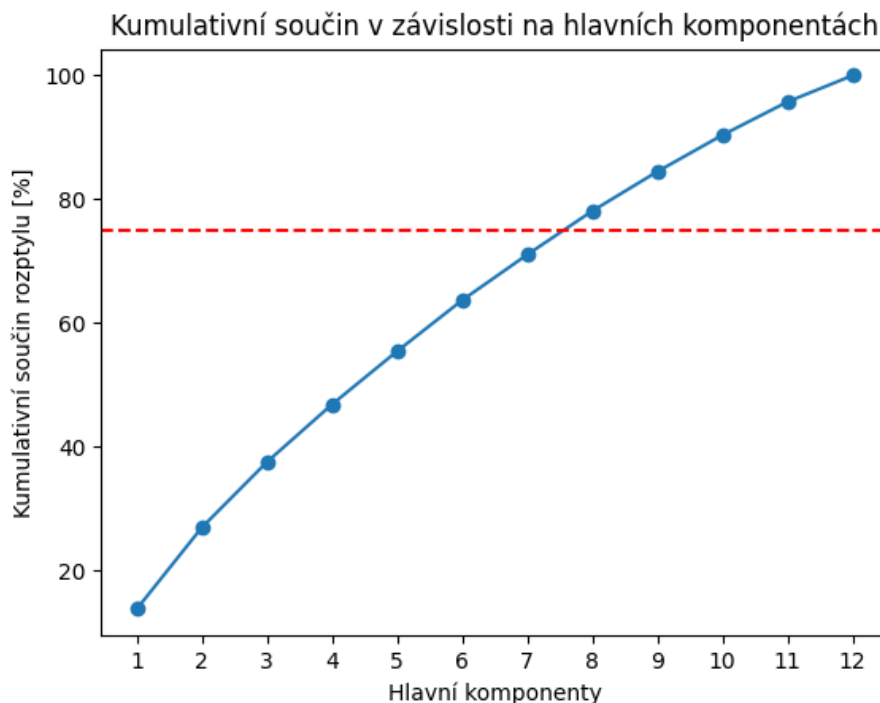
Obr. 4.5: Histogram maximální tepové frekvence [tepy za minutu]



Obr. 4.6: Histogram klidového systolického krevního tlaku [mmHg]

5 Selekcce příznaků pomocí klasického postupu - PCA

PCA bylo realizováno balíčkem scikit-learn (v1.2.2, [75]) v prostředí Python. Každá hlavní komponenta je nová proměnná, která vznikla lineární kombinací původních příznaků za účelem snížit dimenzionalitu. Pro hlavní komponenty je potřeba nejdříve spočítat vlastní čísla a vektory, které byly počítány z korelační matice normalizovaných dat. Jednou z možností jak určit počet hlavních komponent pro klasifikaci je empirický přístup. Vhodnějším, mnohdy i rychlejším, způsobem je z vlastních čísel vypočítat kumulativní součin. Příklad nárůstu rozptylu s rostoucím počtem hlavních komponent je zobrazen na grafu 5.1. Jako práh pro kumulativní součin byla nastavena hodnota 75, což znamená, že vybíráme tolik komponent, aby byl zachován 75% podíl variability (informační hodnoty) z původní datové sady. Na základě tohoto prahu bylo vybráno pro Klevelandský dataset 10 komponent, pro dataset srdečního selhání 8 komponent a pro Indický dataset 11 komponent. Na základě těchto počtů hlavních komponent byla data transformována. Jednotlivé příznaky, které obsahují největší informativní rozptyl v dané komponentě u každého datasetu jsou vypsány v tabulce 5.1. Na základě komponent byla provedena klasifikace, srovnání úspěšnosti predikce s klasifikací bez redukce příznaků pro všechny datasety je uvedeno v tabulce 7.2.



Obr. 5.1: Graf vývoje kumulativního součinu v závislosti na hlavních komponentách pro dataset srdečního selhání

Tab. 5.1: Tabulka zastoupení příznaků, pro dané komponenty

Komponenta	Klevelandský dataset	Dataset srdečního selhání	Indický dataset
1	thal_1 thal_2 slope_2	serum_sodium smoking diabetes	chestpain_2 restingelectro_1 chestpain_1
2	slope_1 slope_2 sex	anemia diabetes high_blood_pressure	restingelectro_0 restingelectro_1 chestpain_2
3	cp_2 cp_3 thal_2	high_blood_pressure anemia diabetes	chestpain_0 chestpain_1 slope_2
4	restecg sex fbc	diabetes high_blood_pressure smoking	exerciseangia slope_1 gender
5	cp_4 cp_3 sex	smoking serum_sodium high_blood_pressure	slope_2 chestpain_0 slope_1
6	sex exang thal_2	time age high_blood_pressure	exerciseangia restingelectro_2 slope_0
7	exang cp_4 cp_2	age time ejection_fraction	restingelectro_2 chestpain_1 slope_1
8	fbc ca slope_0	ejection_fraction age serum_creatine	fastingbloodsugar restingelectro_2 chestpain_2
9	slope_0 ca slope_1	X	gender restingelectro_2 oldpeak
10	cp_1 cp_4 slope_0	X	slope_3 noofmajorvessels slope_0
11	X	X	age noofmajorvessels maxheartrate

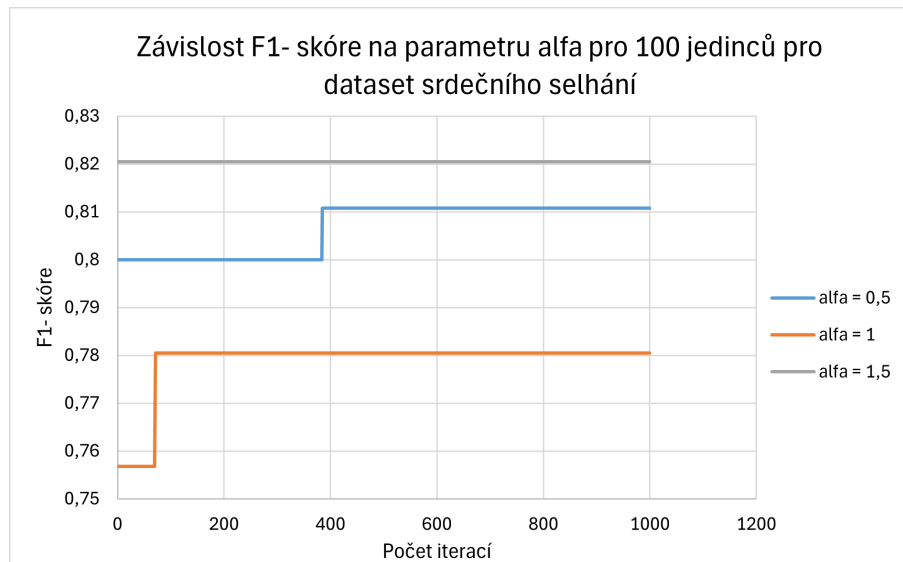
6 Selekcce příznaků meta-heuristickými algoritmy

Pro všechny následující algoritmy platí tyto nastavení. Pro trénování bylo použito 80% dat a pro testování 20% dat. Pro výpočet *fitness* hodnoty byl použit algoritmus RF s počtem 100 rozhodovacích stromů a v každé iteraci se ukládá nejlepší hodnota charakterizující nejlepšího jedince (řešení). Pro každý dataset a všechny kombinace parametrů, které jsou uvedeny níže, bylo spuštěno pět běhů, přičemž do celkového srovnávání byly uvažovány nejlepší běhy na základě *fitness* hodnoty. Počet jedinců byl nastaven na 10, 30, 60 a 100 při ponechání defaultního ukončení 1000 iterací. Další podstatnou úpravou pro tuto úlohu je binarizace, jelikož posuzujeme kvalitu jedince reprezentovaného právě vektorem kombinace nul a jedniček na základě příznaků (0 - nebyl vybrán, 1 - byl vybrán). Jako poslední společná podmínka je zavedena podmínka ošetření klasifikace. Pokud se nám v průběhu iterací v aktuálním jedinci vyskytne pouze vektor nul, je náhodně vybrána jedna pozice, na kterou je umístěna jednička. Toto ošetření zabrání spadnutí algoritmu, ke kterému by došlo, pokud do klasifikátoru pro *fitness* hodnotu nebyl vybrán žádný příznak (nebylo by co klasifikovat). Jako klasifikátory byly využity K-NN s parametrem nejbližních sousedů $k = 3$ a LR. Oba klasifikátory společně s RF byly realizovány balíčkem scikit-learn (v1.2.2, [75]) v prostředí Python.

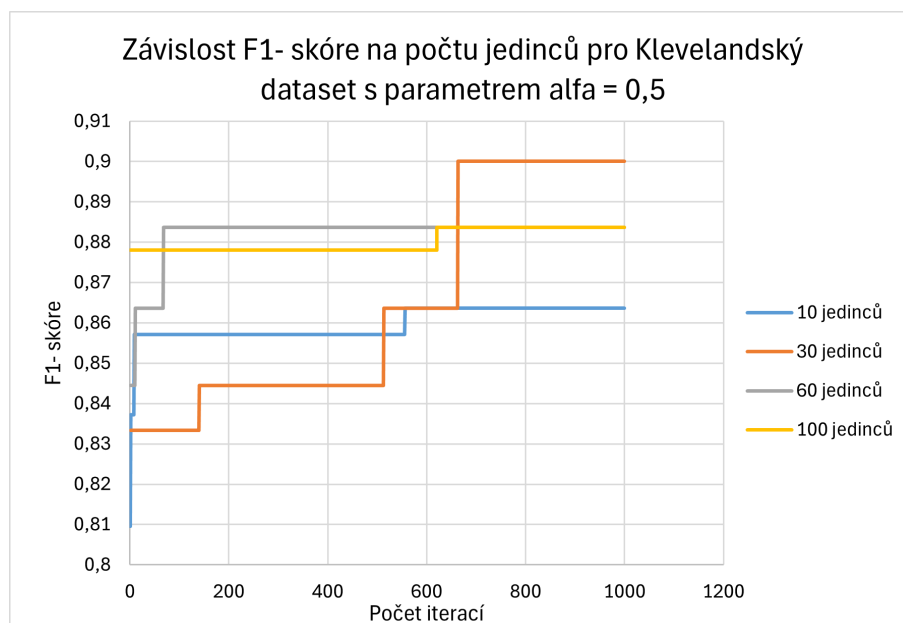
6.1 Selekcce příznaků algoritmem papouška

Počáteční generace je náhodně zvolena podle počtu jedinců a počtu příznaků, které jsou ekvivalentem dimenzí prohledávacího prostoru řešení. Jak již bylo zmíněno v teorii, papoušek může nabývat čtyř stavů. Přičemž ve stavu krmení a sezení na chovateli se uplatňuje krok z pravděpodobnostního rozložení Levyho letu. Pseudokód je znázorněn v příloze 22. Byl vyzkoušen vliv změny parametru alfa v hodnotách 0,5, 1 a 1,5, který ovlivňuje velikost kroků z Levyho letu v prohledávaném prostoru řešení. Srovnání nastavení tohoto parametru pro 100 jedinců pro dataset srdečního selhání je znázorněn na grafu 6.1. Příklad závislosti počtu jedinců u datasetu Kleveland pro přesnost F1- skóre je znázorněna na grafu 6.2. Pro Klevelandský dataset se jevilo nejlepší nastavení 60 a 100 jedinců s krokem alfa = 1. S tímto nastavením vyšla predikce 0,9048, ale vybrané příznaky měly malé odlišnosti. Společně byly vybrány *sex*, *chol*, *restecg*, *thalach*, *ca*, *cp_2* a *slope_2*. Pro dataset srdečního selhání vychází nejlépe nastavení 10 jedinců s parametrem alfa = 1, 30 jedinců s alfa = 1 a alfa = 1,5, 60 jedinců s alfa = 1,5. Pro všechna tato nastavení vyšla *fitness* hodnota 0,8333 na základě příznaků *age*, *ejection_fraction*, *platenes*, *serum_sodium*,

smoking, time. Pro indický dataset se ukázalo nejlepší nastavení populace s 10 jedinci a parametrem $\alpha = 0,5$ a $\alpha = 1,5$, také pro 30 jedinců s $\alpha = 1$ a $\alpha = 1,5$, kdy vyšla fitness hodnota 1. Jako společné příznaky pro tato nastavení byly vybrány *gender, restingBP, serumcholesterol, fastingbloodsugar, slope_3, slope_0, restingelectro_2*.



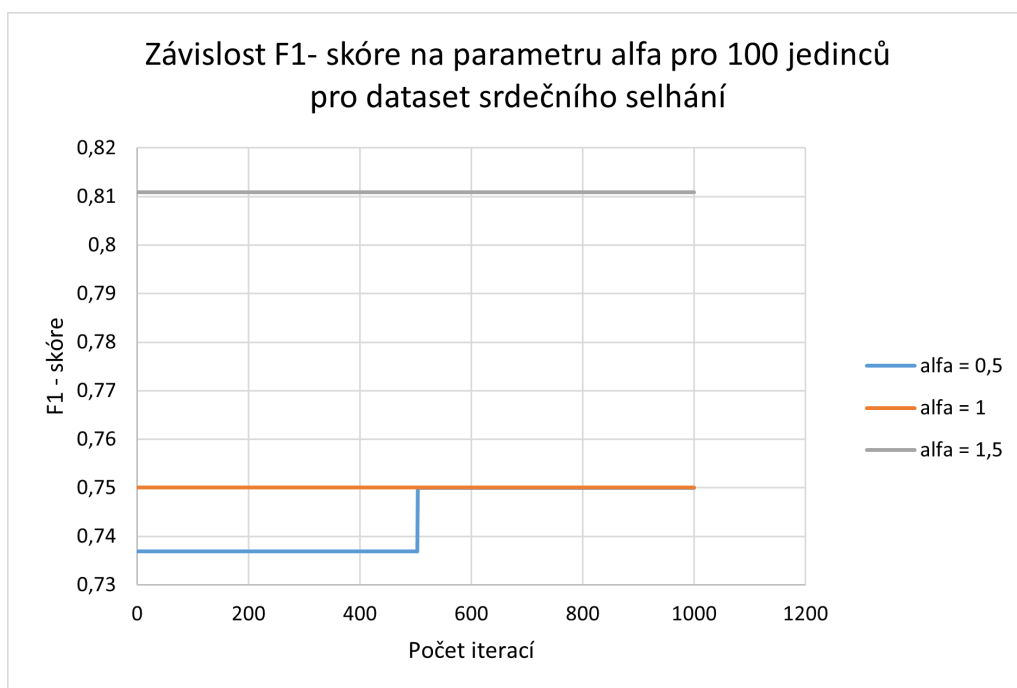
Obr. 6.1: Graf závislosti přesnosti predikce na parametru alfa u datasetu srdečního selhání



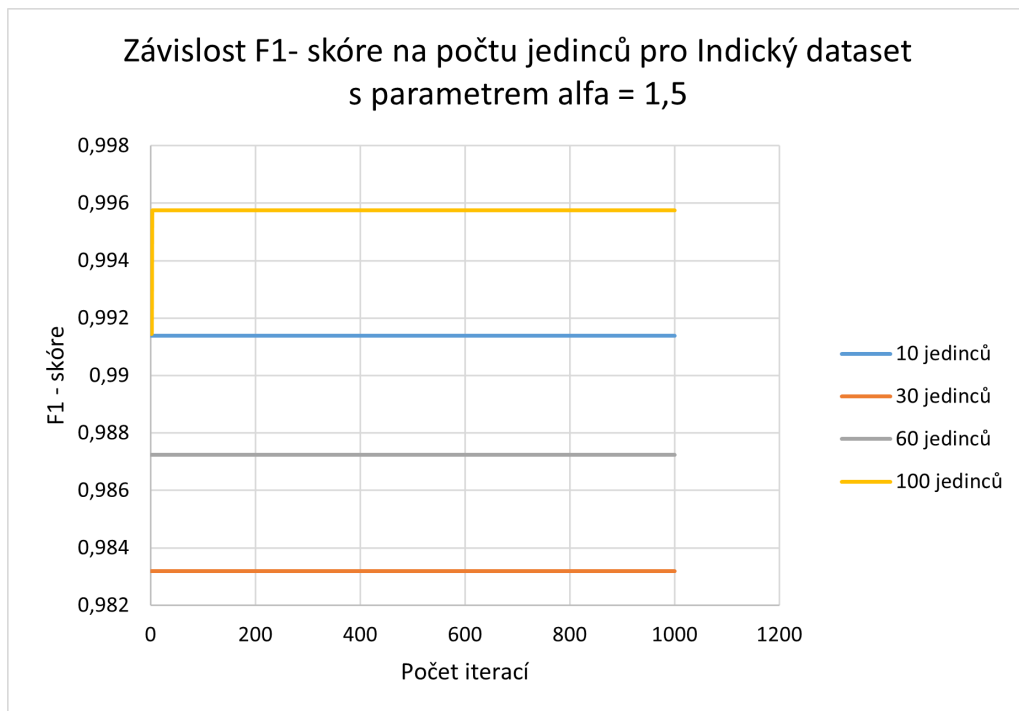
Obr. 6.2: Graf závislosti přesnosti predikce na počtu jedinců pro Klevelandský dataset s parametrem $\alpha = 0,5$

6.2 Selekcce příznaků algoritmem jestřábů

Náhodná počáteční populace je sestavena jako matice počtu jedinců snažících se ulovit kořist a počtu příznaků v datasetu. Obecný postup je znázorněn v příloze 25. K aktualizaci pozic jedinců, při nezdařilém lovu králíka se využívá krok z Levyho letu. Bylo testováno nastavení tohoto parametru na 0,5, 1 a 1,5 s vlivem na výběr příznaků a celkovou přesnost predikce. Srovnání nastavení parametru alfa pro 100 jedinců pro dataset Srdečního selhání je znázorněn na grafu 6.3. Příklad závislosti F1- skóre na volbě počtu jedinců u datasetu Kleveland je znázorněn na grafu 6.4. Pro Klevelandský dataset sejevilo nejlepší nastavení 30 a 60 jedinců s parametrem alfa = 1 a 60 a 100 jedinců s krokem alfa = 1,5. S tímto nastavením vyšla predikce 0,8372. Jako společné byly vybrány příznaky *age*, *sex*, *fbc*, *ca*, *cp_2*. Pro dataset srdečního selhání vychází nejlépe nastavení 60 jedinců s parametrem alfa = 1,5 s fitness hodnotou 0,8544 na základě příznaků *age*, *ejection_fraction*, *platenets*, *serum_sodium*, *smoking*, *time*. Pro indický dataset se ukázalo nejlepší nastavení populace 10 jedinců s alfa = 1 a 100 jedinců a parametrem alfa = 0,5 s fitness hodnotou 0,9957. Jako společné příznaky pro toto nastavení byly vybrány *restingBP*, *fastingbloodsugar*, *maxheartrate*, *chestpain_0*, *chestpain_1*, *slope_1*, *restingelectro_2*.



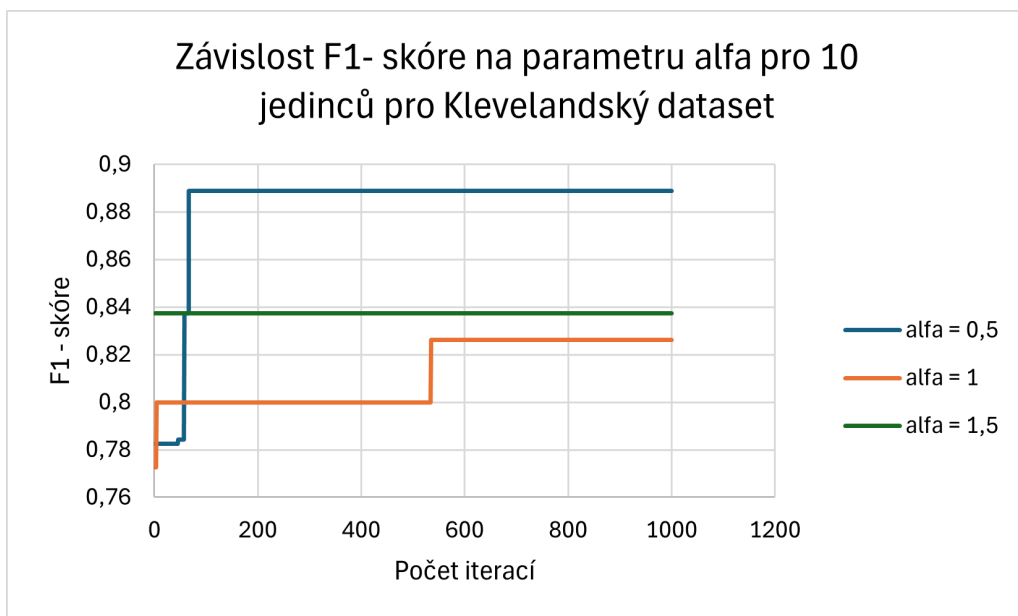
Obr. 6.3: Graf závislosti přesnosti predikce na parametrem alfa pro dataset srdečního selhání



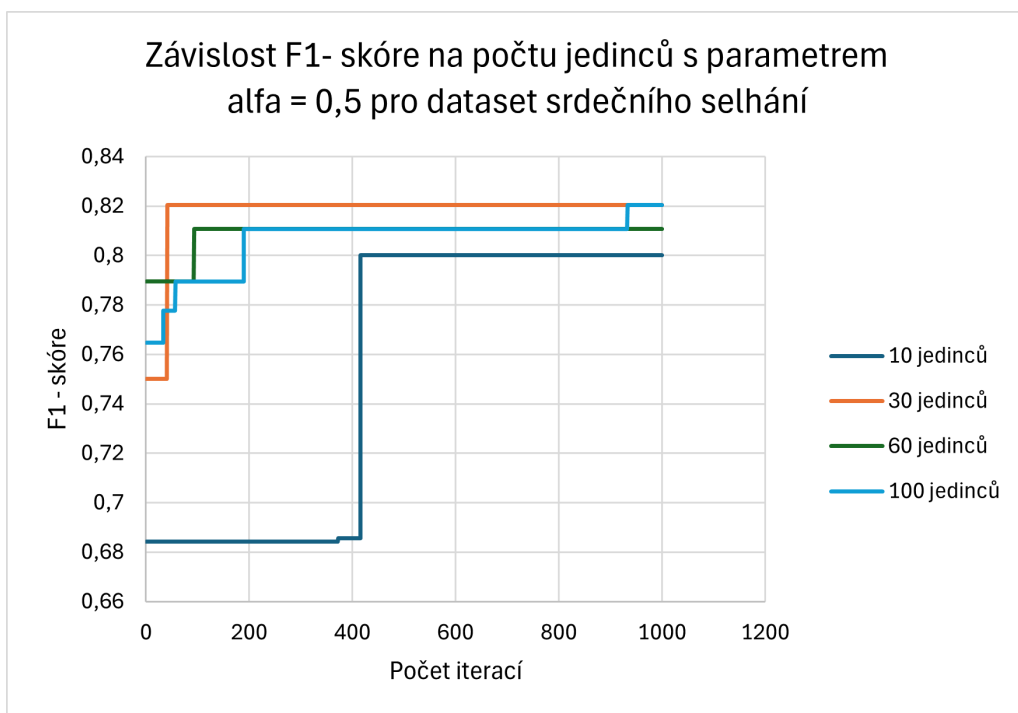
Obr. 6.4: Graf závislosti přesnosti predikce na počtu jedinců s parametru alfa = 1,5 u Indického datasetu

6.3 Selektce příznaků optimalizací opylování květin

Počáteční populace květin je sestavena jako matice závisující na počtu jedinců a počtu příznaků zadaných hodnotami 0 a 1. Obecný postup je znázorněn v příloze 15. Byl testován vliv alfa ovlivňující poletování pylového zrna. Parametr byl nastaven na hodnoty 0,5, 1 a 1,5, příklad vlivu na fitness hodnotu je znázorněn na grafu 6.5. Přesnost predikce se odvíjí i od počtu jedinců, což je znázorněno na grafu 6.6. Pro Klevelandský dataset byla nejlepší fitness hodnota 0,9268 s nastavením 100 jedinců s parametrem alfa = 0,5. Jako relevantní byly vybrány příznaky *sex*, *chol*, *fbc*, *restecg*, *thalach*, *exang*, *oldpeak*, *ca*, *cp_2*, *slope_1*, *slope_0*, *thal_1*, *thal_0*. Pro dataset srdečního selhání vychází nejlépe nastavení 60 jedinců s parametrem alfa = 1 s fitness hodnotou 0,8333 na základě příznaků *age*, *ejection_fraction*, *platenets*, *serum_sodium*, *smoking*, *time*. Pro indický dataset vyšla fitness hodnota 1 při nastavení 10 jedinců s alfa = 0,5 a alfa = 1, 30 jedinců s alfa = 0,5 a 100 jedinců s alfa = 0,5, alfa = 1 a alfa = 1,5. Jako společné příznaky pro tato nastavení byly vybrány *gender*, *restingBP*, *serumcholesterol*, *fastingbloodsugar*, *maxheartrate*, *slope_3*, *restingrelectro_0*.



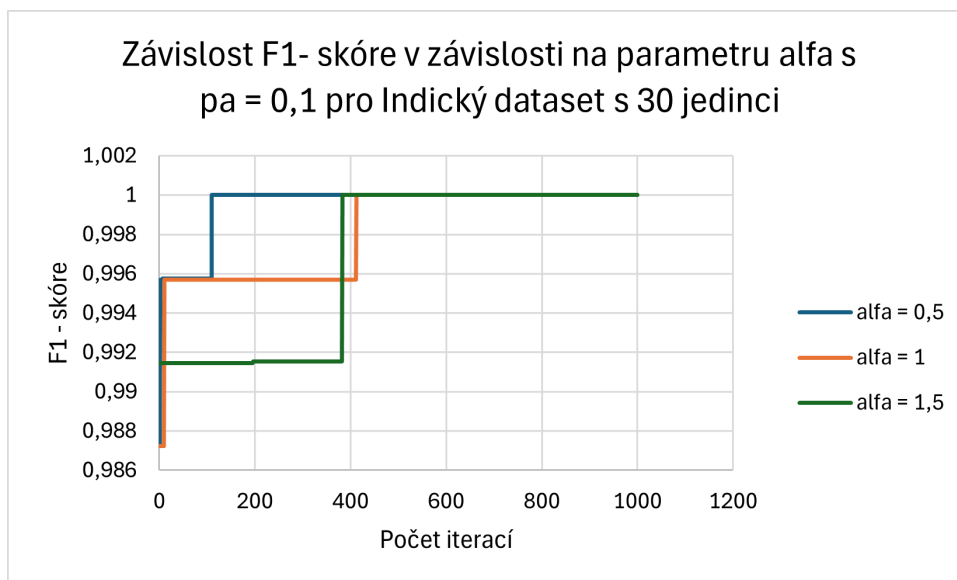
Obr. 6.5: Graf závislosti přesnosti predikce na parametru alfa pro Klevelandský dataset



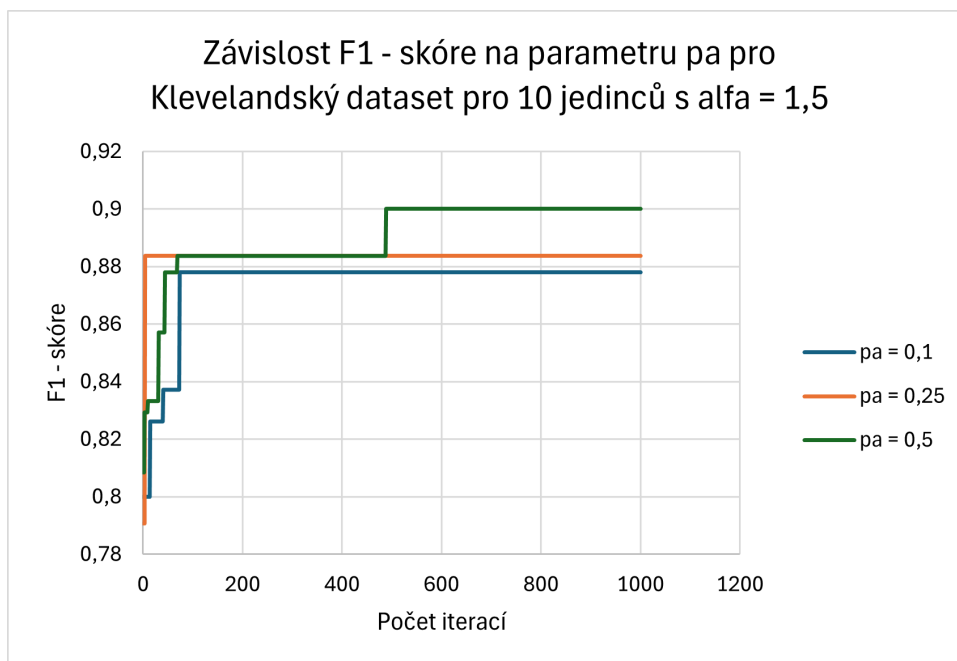
Obr. 6.6: Graf závislosti přesnosti predikce na počtu jedinců s alfa = 0,5 pro dataset srdečního selhání

6.4 Selekcce příznaků pomocí algoritmu kukaček

Počáteční populace je ekvivalentem postavených hnízd s vejci, která odpovídají kandidátům na nejlepší řešení. Obecný postup je znázorněn v příloze 15. Pohyb kukaček, které snáší po jednom vejci do vybraných hnízd je ovlivňován Levyho letem, pro který byl nastaven parametr distribuce alfa na 0, 0,5, 1,5. Ukázkový příklad vlivu nastavení alfa k nalezení nejlepšího řešení je zobrazen na grafu 6.7. Dále byl měněn parametr p_a v hodnotách 0,1, 0,25 a 0,5, který ovlivňuje s jakou pravděpodobností dojde k nalezení parazitního vejce, opuštění stávajícího hnízda (řešení) a postavení nového. Při hodnotě $p_a = 0,5$ je tendence dosahovat lehce horších výsledků. Příklad vývoje fitness hodnoty na základě změny parametru p_a je zobrazen na grafu 6.8. Pro Klevelandský dataset byla nejlepší fitness hodnota 0,9268 při nastavením 60 jedinců s parametrem alfa = 0,5, $p_a = 0,1$ a příznaky *sex*, *chol*, *fbc*, *restecg*, *thalach*, *exang*, *oldpeak*, *ca*, *cp_2*, *slope_1*, *slope_0*, *thal_1*. Pro dataset srdečního selhání je nejlepší hodnota fitness 0,8333 při nastavení alfa = 0,5, $p_a = 0,1$ a $p_a = 0,25$, alfa = 1 pro všechny tři hodnoty p_a , alfa = 1,5 s $p_a = 0,25$ a $p_a = 0,5$ bez ohledu na počet jedinců. Jako nejvíce vypovídající příznaky byly vybrány *age*, *ejection_fraction*, *platenets*, *serum_sodium*, *smoking*, *time*. Pro Indický dataset vyšla fitness hodnota 1 při nastavení alfa = 0,5 se všemi hodnotami p_a bez ohledu na počet jedinců, alfa = 1 pro všechna p_a u 10, 60 a 100 jedinců, alfa = 1 s $p_a = 0,1$ a $p_a = 0,25$ pro 30 jedinců, alfa = 1,5 s $p_a = 0,1$ a $p_a = 0,25$ pro 10 a 30 jedinců, alfa = 1,5 s $p_a = 0,25$ pro 100 jedinců. Jako společné příznaky pro tato nastavení byly vybrány *restingBP*, *serumcholesterol* a *restingelectro_2*.



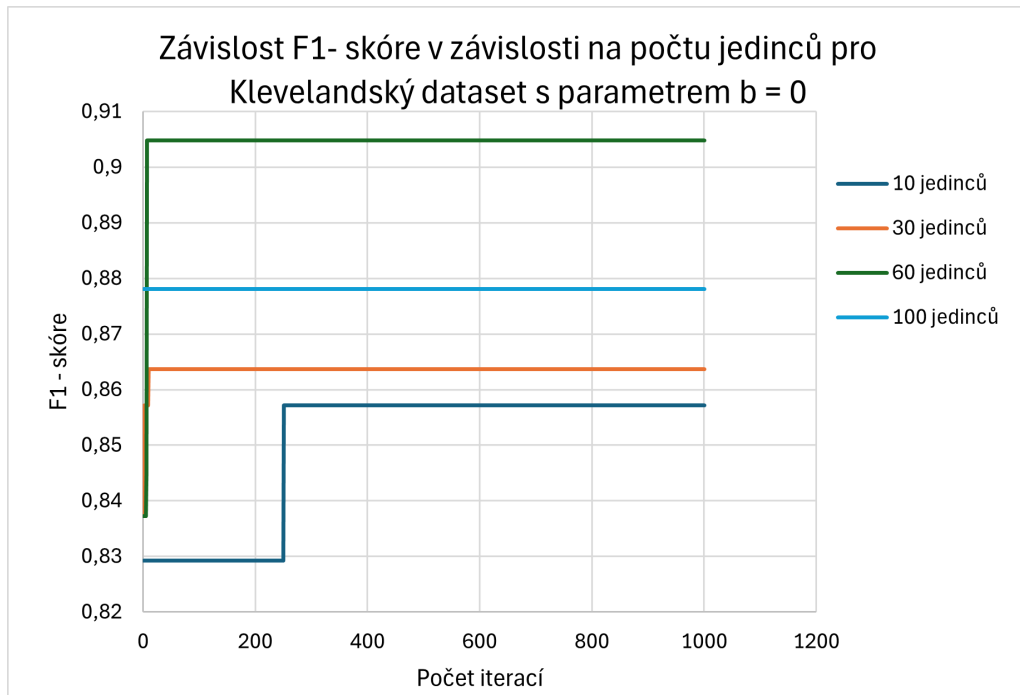
Obr. 6.7: Graf závislosti F1- skóre na nastavení parametru alfa při $p_a = 0,1$ pro Indický dataset



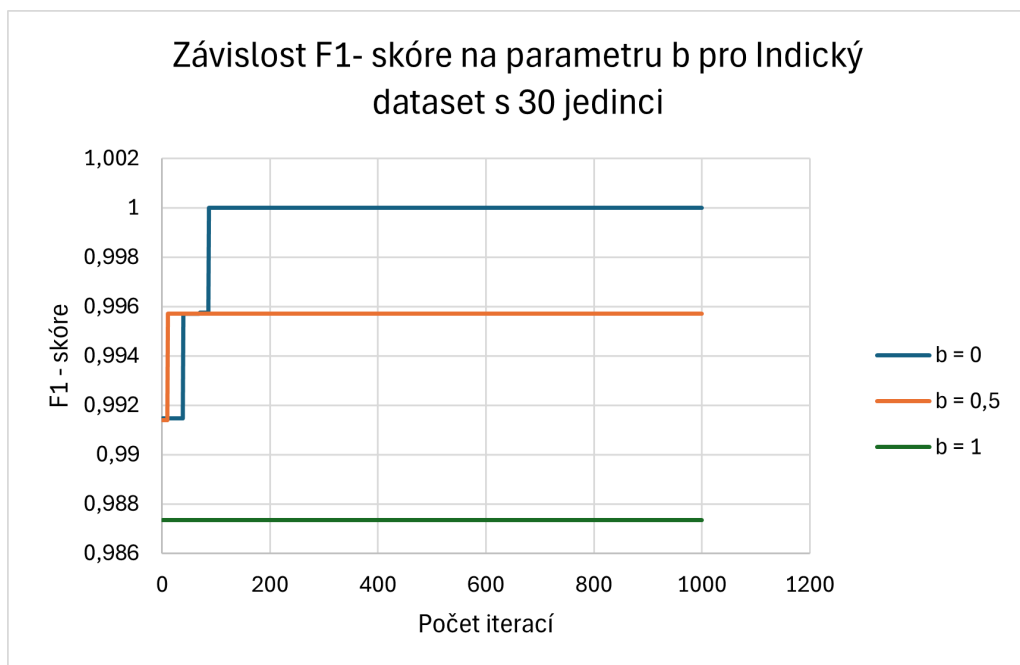
Obr. 6.8: Graf závislosti F1- skóre na parametru p_a při konstantním $\alpha = 1,5$ pro Klevelandský dataset

6.5 Selekcce příznaků optimalizací velryb

Náhodná počáteční populace je sestavena podle počtu jedinců a počtu příznaků v datasetu. Obecný postup je znázorněn v příloze 20. V algoritmu WOA je nastavitelným parametrem beta b , který definuje tvar logaritmické spirály, po které velryby krouží k obklíčení a ulovení kořisti. Tento parametr byl nastaven na hodnoty 0, 0,5 a 1, kdy při hodnotě 0 je spirála menší, nebo-li odpovídá více lokálnímu prohledávání naopak při přiblížení k 1 je větší volnost spirály. Znázornění vlivu beta na fitness hodnotu je vyobrazeno na grafu 6.9. Vliv počtu jedinců při konstantním parametru b na výslednou fitness hodnotu je naznačen v grafu 6.10. Pro Klevelandský dataset byla nejlepší fitness hodnota 0,9048 při nastavení 60 jedinců s parametrem b na hodnotu 0 a příznaky *sex*, *chol*, *restecg*, *thalach*, *oldpeak*, *ca*, *cp_1*, *cp_4*, *cp_2*, *slope_2*, *slope_1*, *thal_1*, *thal_0*, *thal_2*. Pro dataset srdečního selhání je nejlepší hodnota fitness 0,8205 opět při nastavení $b = 0$ na základě příznaků *age*, *ejection_fraction*, *high_blood_pressure*, *platenets*, *serum_sodium*, *time*. Pro Indický dataset vyšla fitness hodnota 1 při nastavení $b = 0,5$ a $b = 1$ pro 10 jedinců, $b = 0$ pro 30 a 60 jedinců a $b = 1$ pro 100 jedinců. Jako společné příznaky pro tato nastavení byly vybrány *gender*, *restingBP*, *serumcholesterol*, *fastingbloodpressure* a *slope_1*.



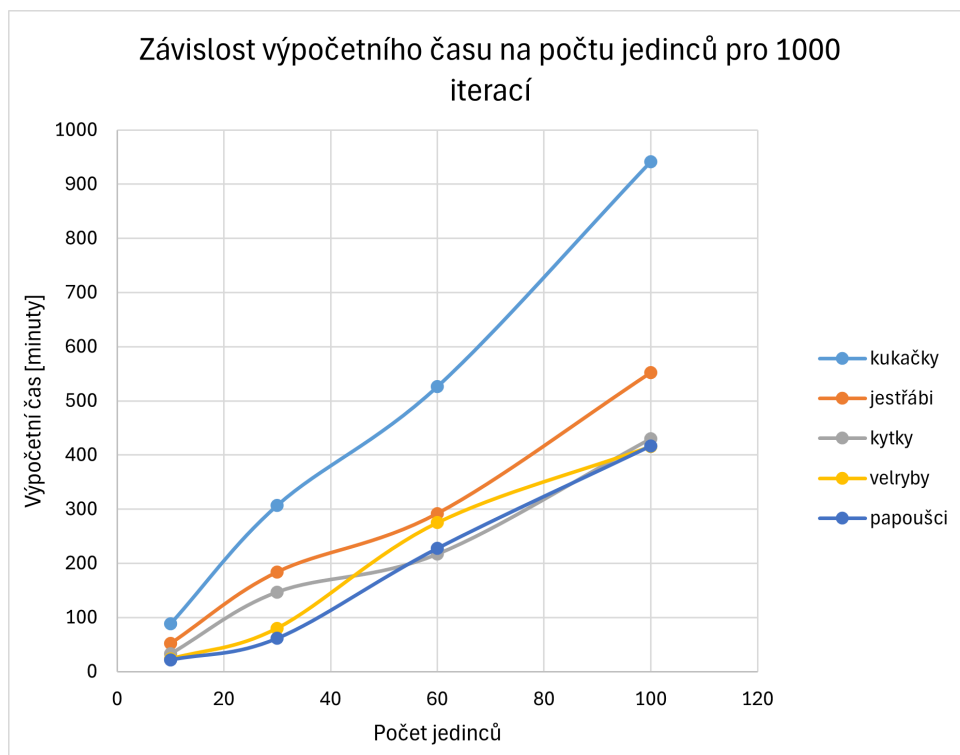
Obr. 6.9: Graf závislosti F1- skóre na počtu jedinců při konstantní b pro Klevelandský dataset



Obr. 6.10: Graf závislosti F1- skóre na změně parametru b u Indického datasetu

7 Srovnání klasifikace na základě vybraných příznaků všemi přístupy

Kompletní výsledky pro kombinace testovaných parametrů jsou uvedeny v příloze E. Nejlepší volby algoritmů s parametry pro jednotlivé datasety jsou shrnuty v tabulce 7.3. Na základě použitého klasifikačního algoritmu dosahuje K-NN horších výsledků než použití LR a nejlépe klasifikuje RF, který byl využit pro výpočet *fitness* hodnoty. Oproti použití všech příznaků a běžnému postupu PCA dosahují meta-heuristické algoritmy lepší predikce. Nelze však na základě výsledků s určitostí říci, který z použitých meta-heuristických algoritmů dosahuje obecně nejlepších výsledků, jelikož predikce závisí převážně na hodnotách příznaků v datasetech. Dále je také nutné uvažovat vysokou výpočetní náročnost z důvodu opakovaného výpočtu *fitness* hodnoty při změnách v populaci. Srovnání závislosti časové náročnosti na počtu jedinců pro 1000 iterací na datasetu Kleveland je ukázáno na grafu 7.1. Nastává také potřeba zvážit kompromis mezi počtem jedinců, počtem iterací a celkovou úspěšností predikce. Časově nejnáročnější se ukazuje CSO, jelikož dochází nejčastěji k aktualizaci jedinců na základě Levyho letu a stavbě nových hnízd (nová možná řešení) a tudíž je velice častý krok ohodnocení kvality populace. Závislosti F1- skóre (*fitness* hodnota) na počtu jedinců a nastavených parametrech pro všechny meta-heuristické algoritmy u Klevelandského datasetu jsou zobrazeny na grafech 7.2, 7.3, 7.4, 7.5, 7.6. Dle výsledku největší náchylnost na změnu parametrů a počet jedinců má HHO, zatím co za nejstabilnější řešení při změnách parametrů a jedinců lze považovat CSO. Na základě výsledků můžeme konstatovat, že ne vždy největší počet jedinců znamená nejlepší řešení. Co se týče nastavitelného parametru alfa, obecně čím blíže je k hodnotě 2, tím menší skokový vliv během iterací má na výsledné řešení. U většiny řešení při $\alpha = 1,5$ zůstává počáteční hodnota konstantní a pohyb s velkými kroky se neprojeví na změně nejlepšího jedince.



Obr. 7.1: Graf závislosti výpočetního času na počtu jedinců při 100 iterací pro Klevelandský dataset

Tab. 7.1: Tabulka přesnosti klasifikace na základě všech příznaků

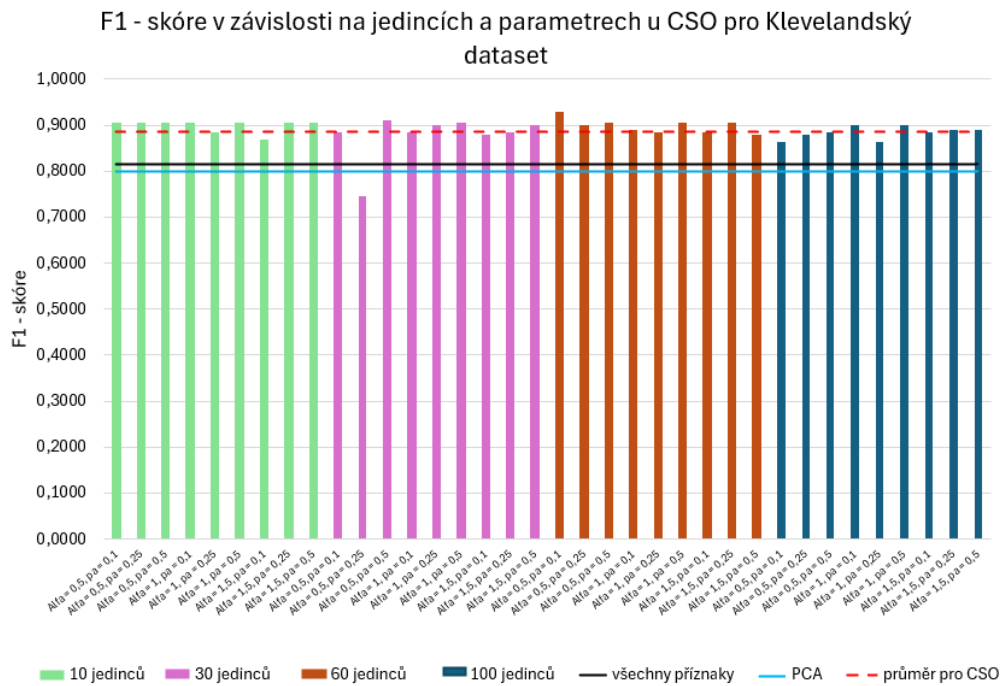
X	Všechny příznaky		
Dataset	F1-skóre	Specifická	Senzitivita
Kleveland	0,8160	0,8610	0,8333
Srdeční selhání	0,6150	0,9433	0,4800
Indie	0,9870	0,9889	0,9833

Tab. 7.2: Tabulka přesnosti klasifikace na základě PCA

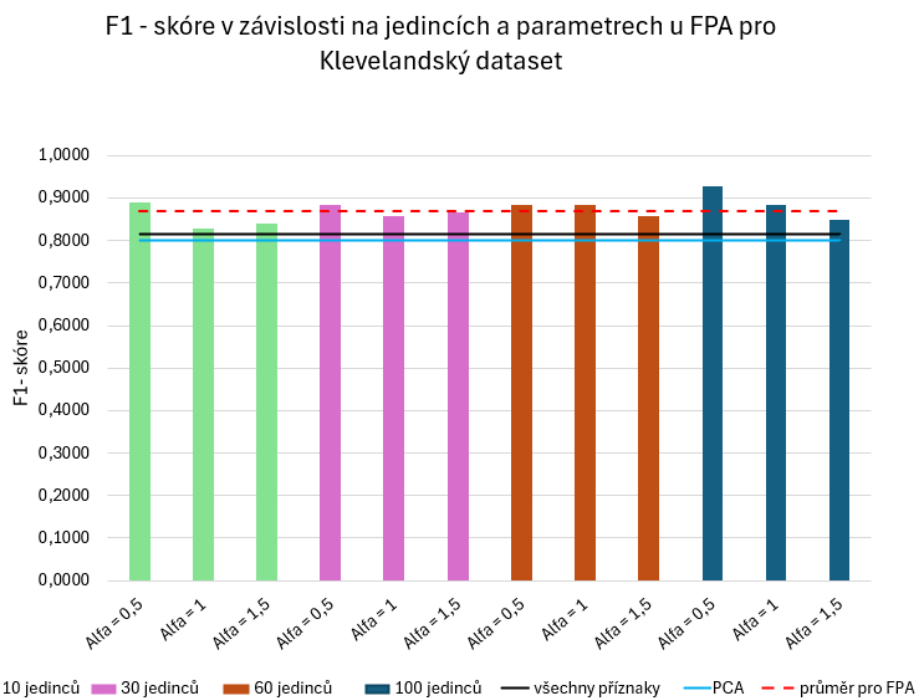
X	PCA		
Dataset	F1-skóre	Specifická	Senzitivita
Kleveland	0,8000	0,8333	0,8333
Srdeční selhání	0,5556	0,9710	0,4000
Indie	0,9700	0,9400	0,9833

Tab. 7.3: Tabulka nejlepších algoritmů na základě F1- skóre

Algoritmus	Jedinci	Parametry	F1 - skóre	Specificita	Senzitivita
Klevelandský dataset					
FPA	100	alfa = 0,5	0,9268	0,8684	0,8182
CSO	60	alfa = 0,5, pa = 0,1	0,9268	0,8684	0,8182
Dataset srdečního selhání					
HHO	60	alfa = 1,5	0,8544	0,8422	0,5238
Indický dataset					
PO	10	alfa = 0,5	1,0000	0,9759	0,9658
	10	alfa = 1,5	1,0000	0,9759	0,9658
	30	alfa = 1	1,0000	0,9639	0,9658
	30	alfa = 1,5	1,0000	0,9879	0,9658
WOA	10	beta = 0,5	1,0000	0,9880	0,9658
	10	beta = 1	1,0000	0,9880	0,9658
	30	beta = 0	1,0000	0,9880	0,9658
	60	beta = 0	1,0000	0,9639	0,9145
	100	beta = 1	1,0000	0,9759	0,9658
FPA	10	alfa = 0,5	1,0000	0,9880	0,9658
	10	alfa = 1	1,0000	0,9880	0,9658
	30	alfa = 0,5	1,0000	0,9759	0,9658
	100	alfa = 0,5	1,0000	0,9759	0,9744
	100	alfa = 1	1,0000	0,9759	0,9658
	100	alfa = 1,5	1,0000	0,9639	0,9829
CSO	10	alfa = 0,5, pa = 0,1	1,0000	0,9157	0,9145
	10	alfa = 0,5, pa = 0,25	1,0000	0,9759	0,9658
	10	alfa = 0,5, pa = 0,5	1,0000	0,9759	0,9829
	10	alfa = 1, pa = 0,1	1,0000	0,9880	0,9744
	10	alfa = 1, pa = 0,25	1,0000	0,9880	0,9658
	10	alfa = 1, pa = 0,5	1,0000	0,9880	0,9573
	10	alfa = 1,5, pa = 0,1	1,0000	0,9880	0,9658
	10	alfa = 1,5, pa = 0,25	1,0000	0,9639	0,9573
	30	alfa = 0,5, pa = 0,1	1,0000	0,9880	0,9658
	30	alfa = 0,5, pa = 0,25	1,0000	0,9759	0,9658
	30	alfa = 0,5, pa = 0,5	1,0000	0,9759	0,9658
	30	alfa = 1, pa = 0,1	1,0000	0,9639	0,9829
	30	alfa = 1, pa = 0,25	1,0000	0,9639	0,9658
	30	alfa = 1,5, pa = 0,1	1,0000	0,9880	0,9658
	30	alfa = 1,5, pa = 0,25	1,0000	0,9639	0,9658
	60	alfa = 0,5, pa = 0,1	1,0000	0,9518	0,9658
	60	alfa = 0,5, pa = 0,25	1,0000	0,9759	0,9573
	60	alfa = 0,5, pa = 0,5	1,0000	0,9759	0,9658
	60	alfa = 1, pa = 0,1	1,0000	0,9759	0,9658
	60	alfa = 1, pa = 0,25	1,0000	0,9880	0,9658
	60	alfa = 1, pa = 0,5	1,0000	0,9880	0,9573
	100	alfa = 0,5, pa = 0,1	1,0000	0,9759	0,9658
	100	alfa = 0,5, pa = 0,25	1,0000	0,9759	0,9658
	100	alfa = 0,5, pa = 0,5	1,0000	0,9759	0,9658
	100	alfa = 1, pa = 0,1	1,0000	0,9639	0,9829
	100	alfa = 1, pa = 0,25	1,0000	0,9880	0,9658
	100	alfa = 1, pa = 0,5	1,0000	0,9880	0,9573
	100	alfa = 1,5, pa = 0,25	1,0000	0,9639	0,9658

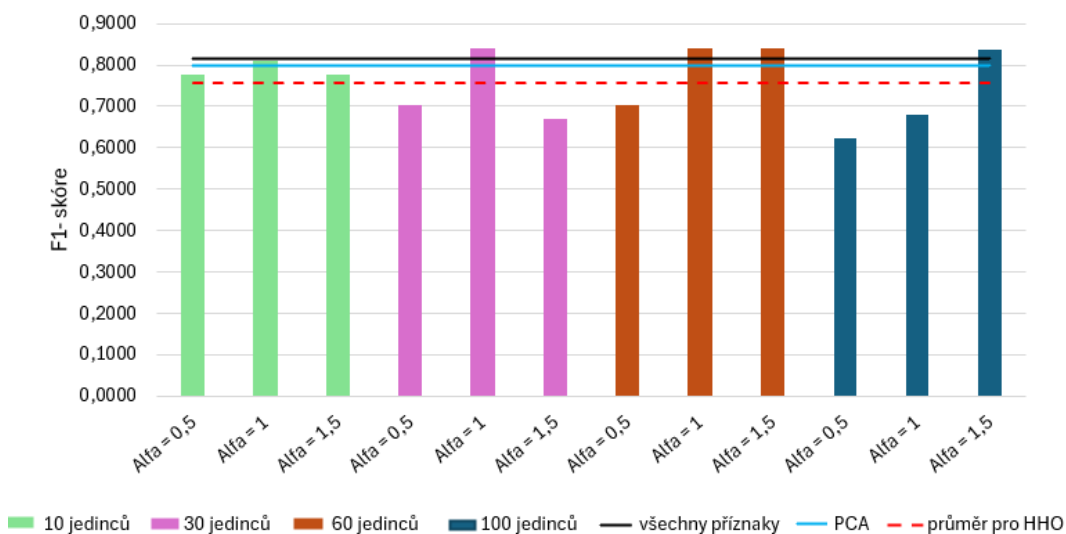


Obr. 7.2: Graf závislosti F1- skóre na počtu jedinců a parametrech u CSO pro Klevelandský dataset



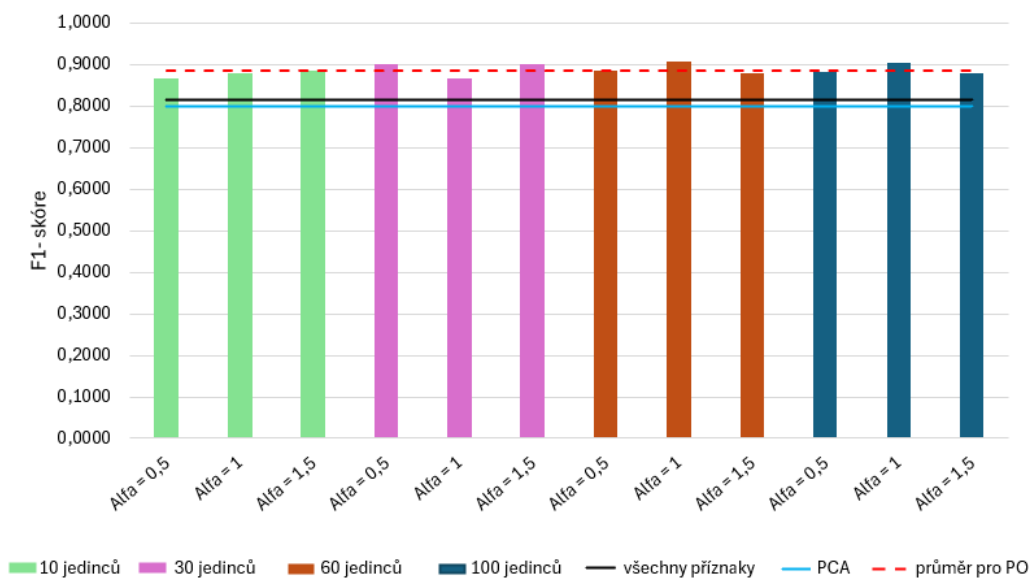
Obr. 7.3: Graf závislosti F1- skóre na počtu jedinců a parametrech u FPA pro Klevelandský dataset

F1 - skóre v závislosti na jedincích a parametrech u HHO pro Klevelandský dataset



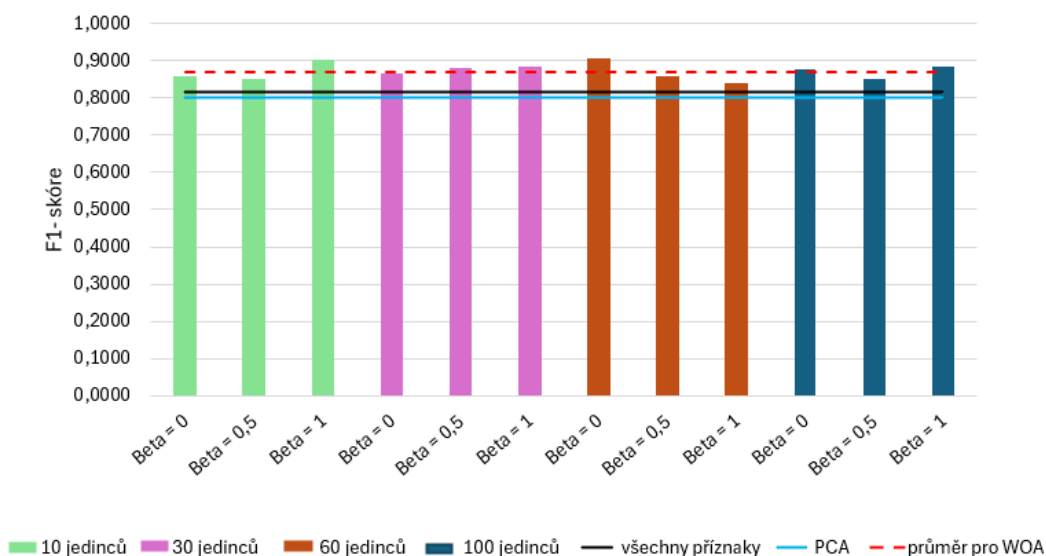
Obr. 7.4: Graf závislosti F1- skóre na počtu jedinců a parametrech u HHO pro Klevelandský dataset

F1 - skóre v závislosti na jedincích a parametrech u PO pro Klevelandský dataset



Obr. 7.5: Graf závislosti F1- skóre na počtu jedinců a parametrech u PO pro Klevelandský dataset

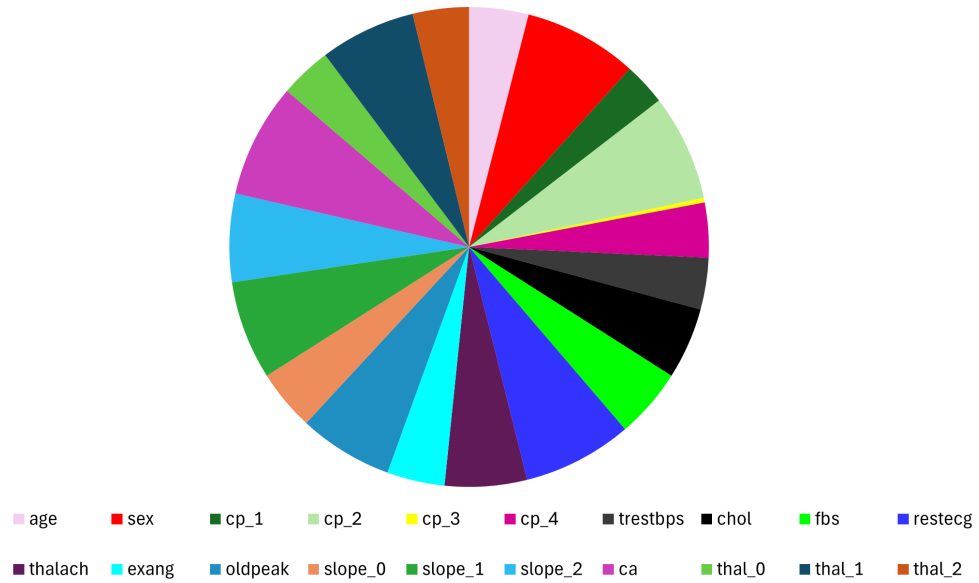
F1 - skóre v závislosti na jedincích a parametrech u WOA pro Klevelandský dataset



Obr. 7.6: Graf závislosti F1- skóre na počtu jedinců a parametrech u WOA pro Klevelandský dataset

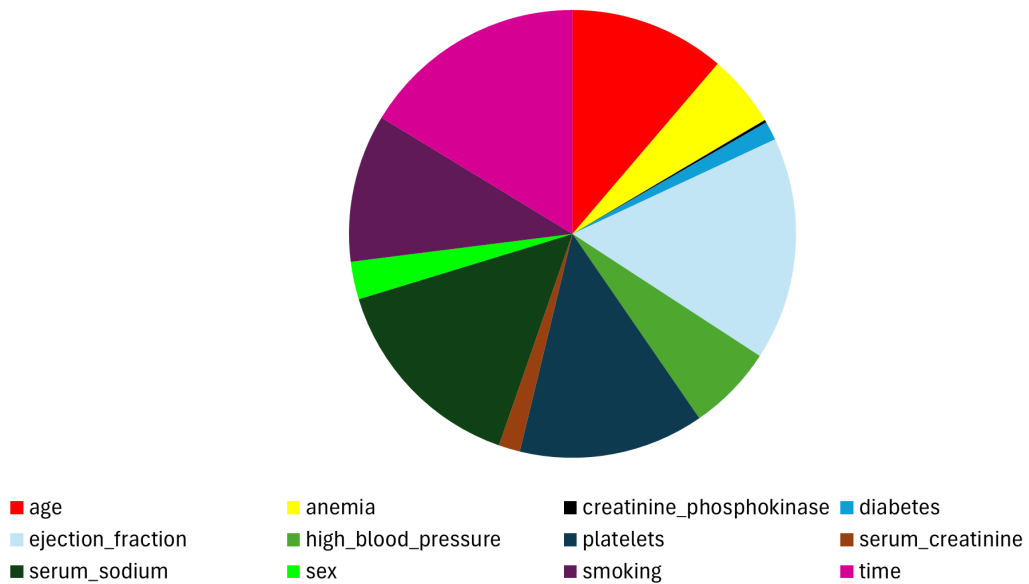
V případě kombinace příznaků a vytvoření souboru vektorů pro nejlepší predikci (nejlepší jedinec) je vysoký možný počet kombinací, u Klevelandského a Indického datasetu můžeme dosáhnout až 1048575 možných kombinací a u datasetu srdečního selhání 4095 kombinací. Všeobecně u všech algoritmů a datasetů byl každý příznak alespoň jednou vybrán jako relevantní pro klasifikaci. Zasloupení celkových vybrání pro všechny meta-heuristické algoritmy se všemi parametry u jednotlivých datasetů je zobrazeno na grafech 7.7, 7.8, 7.9. Pro Klevelandský a Indický dataset je patrné rovnoměrnější zastoupení většiny příznaků, nesou podobnou informační hodnotu pro výslednou klasifikaci, zatím co u datasetu srdečního selhání převažuje výběr pouze některých příznaků. Pro Klevelandský dataset se nejčastěji vyskytovaly příznaky: pohlaví, klidové EKG, atypická bolest na hrudi a počet koronárních tepen postižených stenózou. V případě datasetu srdečního selhání byly nejčastěji vybrány jako relevantní příznaky: objem krve vypuzené srdcem, obsah sodíku v séru, obsah krevních destiček v krvi a sledovací období. Pro Indický dataset byly nejčastěji vybrány příznaky: klidový systolický krevní tlak, hladina cholesterolu v krvi, abnormalita levé komory na EKG, glykémie na lačno, sestupný a vzestupný ST úsek při zátěži.

Zastoupení jednotlivých příznaků podle počtu vybrání meta-heuristickými algoritmy pro Klevelandský dataset



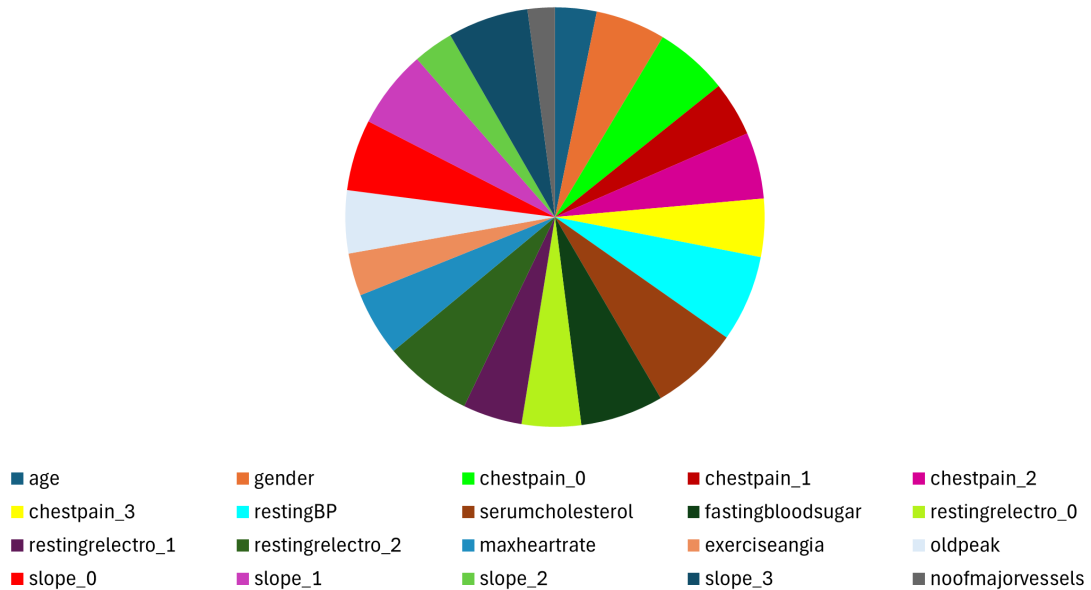
Obr. 7.7: Graf zastoupení počtu vybrání jednotlivých příznaků všemi algoritmy pro Klevelandský dataset

Zastoupení jednotlivých příznaků podle počtu vybrání meta-heuristickými algoritmy pro dataset srdečního selhání



Obr. 7.8: Graf zastoupení počtu vybrání jednotlivých příznaků všemi algoritmy pro dataset srdečního selhání

Zastoupení jednotlivých příznaků podle počtu vybrání meta-heuristickými algoritmy pro Indický dataset



Obr. 7.9: Graf zastoupení počtu vybrání jednotlivých příznaků všemi algoritmy pro Indický dataset

Nejlepší predikce dosahovaly algoritmy v případě Indického datasetu, který obsahuje i nejvíce pacientů. Při testování augmentace trénovacích dat u zbylých dvou datasetů na podobnou velikost trénovací množiny však nebyl zjištěn markantní výsledek ve zlepšení predikce. Na tomto zjištění lze předpokládat více odpovídající hodnoty pro srdeční onemocnění jednotlivých příznaků. Pro dataset srdečního selhání vychází predikce a především senzitivita všeobecně hůře z důvodu jiných příznaků a hlavně se úmrtí či přežití vztahuje na sledovací období. U některých pacientů je sledovací období jen 4 dny a u některých třeba 100, v obou případech pacient přežívá nebo umírá. Tudíž pacient sice může mít špatné markery, které by predikovaly srdeční selhání, ale pokud je pacient sledován velmi krátkou dobu, je možné že po toto období přežil.

Srovnání se studií [1], ze které tato práce vychází je shrnuto v tabulce 7.4. Kvalita klasifikace podle F1- skóre, pro Klevelandský dataset při otestovaných nastaveních, vychází pro CSO u LR průměrně o 7,9% horší, u K-NN o 20% horší, u RF o 4,5% lepší, pro FPA u LR o 5% horší, u K-NN o 22% horší a u RF o 5,9% lepší, pro WOA u LR o 5,1% horší, u K-NN o 20,7% horší a u RF o 1,8% lepší, pro HHO u LR o 16,2% horší, u K-NN o 27% horší a u RF o 10% horší. Většinou jako společné relevantní příznaky ve studii a v této práci byly vybrány pro CSO *sex, restecg, ca, slope_2*, pro FPA *age, ca, thal_2*, pro WOA *sex, ca, thal_2* a pro HHO *sex, restecg,*

ca. Pro dataset srdečního selhání jsou výsledky pro CSO u LR průměrně o 4,9% horší, u K-NN o 26,3% horší, u RF o 9,5% lepší, pro FPA u LR o 9% horší, u K-NN o 43% horší a u RF o 4,7% lepší, pro WOA u LR o 5,1% horší, u K-NN o 42,4% horší a u RF o 9,3% lepší, pro HHO u LR o 4,8% horší, u K-NN o 24,4% horší a u RF o 6,6% lepší. Většinou jako společné relevantní příznaky ve studii a v této práci byly vybrány pro CSO *ejection_fraction*, *time*, pro FPA *ejection_fraction*, *high_blood_pressure*, pro WOA *time* a pro HHO *time*.

Tab. 7.4: Souhrnná tabulka srovnání výsledků studie a této práce na základě průměrných hodnot pro všechna nastavení daných algoritmů

Klevelandský dataset						
X	Výsledky studie [1]			Výsledky této práce		
Algoritmus	RF	LR	KNN	RF	LR	KNN
CSO	0,8438	0,8729	0,9794	0,8890	0,7938	0,7498
FPA	0,8111	0,8647	0,9662	0,8701	0,8145	0,7455
WOA	0,8522	0,8522	0,9480	0,8703	0,8014	0,7428
HHO	0,8582	0,8650	0,9587	0,7566	0,7027	0,6884
Dataset srdečního selhání						
Algoritmus	RF	LR	KNN	RF	LR	KNN
CSO	0,7351	0,7548	0,9534	0,8305	0,7061	0,6907
FPA	0,7567	0,7319	0,9558	0,8037	0,6412	0,5235
WOA	0,7068	0,6846	0,9427	0,7992	0,6359	0,5180
HHO	0,7228	0,6901	0,9348	0,7892	0,6467	0,6909

Porovnání s některými zveřejněnými studiemi k těmto datasetům je shrnuté v tabulce 7.5. Meta-heuristické algoritmy ve většině případů vycházejí s lepší predikcí. Pro Klevelandský dataset se jeví jako vhodný algoritmus FPA, pro dataset srdečního selhání je vhodné použití HHO a pro Indický dataset vyšly nejlépe algoritmy z této předkládané práce. Konkrétně algoritmus kukaček, velryb, papouška a opylování květin v kombinaci s klasifikátorem náhodné doménové struktury .

Tab. 7.5: Tabulka srovnání výsledků dostupných studií a této práce na základě hodnot F1- skóre

Studie (literatura)	Rok	Model	F1 - skóre
Klevelandský dataset			
Patro et al. [76]	2022	Sestava váhovaných klasifikátorů založená na přesnosti (AB-WAE)	0,9300
Ambrish et al. [77]	2022	LR	0,8710
Zhenya and Zhang [78]	2021	Hybridní sestava klasifikátorů (Ensemble)	0,8880
Hera et al. [79]	2022	Více - úroňňová predikce (Multi-Tier Ensemble - MTE)	0,7914
Studie [1]	2022	FPA - KNN	0,9972
Tato práce	2024	FPA - RF, CSO - RF	0,9268
Dataset srdečního selhání			
Chicco and Jurman [80]	2020	Uspořádnání příznaků + LR	0,7190
Ishaq et al. [81]	2021	Nadměrné vzorkování syntetických menšin (SMOTE) + RF	0,9188
Nishat et al. [82]	2022	Nadměrné vzorkování syntetických menšin(SMOTE) + LR	0,8840
Karakuş and Er [83]	2022	Metoda podpůrných vektorů s gaussovským jádrem (Gaussian SVM)	1,0000
Studie [1]	2022	HHO - KNN	0,9745
Tato práce	2024	HHO - RF	0,8544
Indický dataset			
Doppala et al. [84]	2022	Posilování gradientu (XGBoost)	0,9612
Anu and David [85]	2021	Rozhodovací strom (DP) + CatBoost (CT) + LR	0,9880
Tato práce	2024	WOA - RF, FPA - RF, CSO - RF PO - RF	1,0000

Závěr

Náplní předkládané diplomové práce bylo seznámit čtenáře s přírodou inspirovanými meta-heuristickými algoritmy a obecným postupem při klasifikační úloze. Praktická část aplikuje teoreticky popsané postupy na Klevelandském datasetu, datasetu srdečního selhání a Indickém datasetu. Přesnost predikce srdečního onemocnění byla vyhodnocena pomocí algoritmů LR, KNN a RF na základě vybraných příznaků pomocí PCA a různých metaheuristických algoritmů s různým počtem jedinců a kombinací nastavitelných parametrů. Nebyly zjištěny markantní rozdíly v různých nastaveních a kombinacích parametrů. Nelze však průkazně označit některý z algoritmů za nejlepší, lze však konstatovat dosažení lepších výsledků výběrem příznaků meta-heuristickými algoritmy než běžně používanými metodami. Nutné je však zvážit časovou náročnost algoritmů, která je na těchto datasetech násobně větší oproti klasickým metodám. Na základě výsledků by nebylo nutné ponechat pro všechny algoritmy defaultní nastavení 1000 iterací. U většiny algoritmů by stačilo jen 500 iterací případně nastavit kritérium dosažení určitého F1- skóre, jelikož v závislosti na náhodě vstupující do počáteční populace a rozložení na trénovací a testovací data už první jedinec (možné řešení) dosahuje vysoké kvality. Pro dataset srdečního selhání vychází predikce a především senzitivita všeobecně hůře pravděpodobně z důvodu jiných příznaků a úmrtí či přežití se vztahuje na sledovací období, které není pro všechny pacienty stejné. Pro Klevelandský dataset se nejčastěji vyskytovaly příznaky: pohlaví, klidové EKG a počet koronárních tepen postižených stenózou. V případě datasetu srdečního selhání byly nejčastěji vybrány jako relevantní příznaky: objem krve vypuzené srdcem, obsah sodíku v séru a sledování období. Pro Indický dataset byly nejčastěji vybrány příznaky: klidový systolický krevní tlak, hladina cholesterolu v krvi a abnormalita levé komory u EKG.

Předložená práce ukazuje, že postup výběru příznaků pro úlohu klasifikace je řešitelný i s využitím evolučních algoritmů. V porovnání se standardními metodami redukce příznaků dosahují všechny použité metody lepších výsledků na všech třech použitých datových sadách. V porovnání s odbornou literaturou jsou některé výsledky mírně horší.

Literatura

- [1] AY, Şevket; GARIP, Zeynep a EKINCI, Ekin, 2023. *A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases*. Online. The Journal of Supercomputing. 79, s. 11797–11826. DOI: <https://doi.org/10.1007/s11227-023-05132-3>. [cit. 2023-12-31].
- [2] ABIODUN, Esther Omolara; ALABDULATIF, Abdulatif; ABIODUN, Oludare Isaac; ALAWIDA, Moatsum a ALKHAWALDEH, Rami S., 2021. *A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities*. Online. Neural Computing and Applications. 33, s. 15091–15118. DOI: <https://doi.org/10.1007/s00521-021-06406-8>. [cit. 2023-12-28].
- [3] GARCÍA-RODRIGUEZ, José; SALVADOR-MENESES, Jaime a RUIZ-CHAVEZ, Zoila, 2018. *Machine Learning Methods Based Preprocessing to Improve Categorical Data Classification*. Online. In: Intelligent Data Engineering and Automated Learning – IDEAL 2018. Springer International Publishing, s. 297-304. ISBN 978-3-030-03493-1. Dostupné z: https://link.springer.com/chapter/10.1007/978-3-030-03493-1_32. [cit. 2023-11-22].
- [4] ILIOU, Theodoros; ANAGNOSTOPOULOS, Christos-Nikolaos; ANASTASOPOULOS, George a NERANTZAKI, Marina, 2015. *A Novel Machine Learning Data Preprocessing Method for Enhancing Classification Algorithms Performance*. Online. 2015. DOI: <https://doi.org/10.1145/2797143.2797155>. [cit. 2023-11-22].
- [5] SINGH, Dalwinder a SINGH, Birmohan, 2020. *Investigating the impact of data normalization on classification performance*. Online. Applied Soft Computing. 97, s. 105524. ISSN 1568-4946. DOI: <https://doi.org/10.1016/j.asoc.2019.105524>. [cit. 2023-11-22].
- [6] GAL, Michal a RUBINFELD, Daniel, 2018. *Data Standardization*. Online. SSRN Electronic Journal. ISSN 1556-5068. DOI: <https://doi.org/10.2139/ssrn.3326377>. [cit. 2023-11-22].
- [7] JAADI, Zakaria. *When and Why to Standardize Your Data*. Online. In: BuiltIn. Dostupné z: <https://builtin.com/data-science/when-and-why-standardize-your-data>. [cit. 2023-11-22].

- [8] *Classification in Machine Learning: An Introduction*, 2022. Online. In: Data camp. Dostupné z: <https://www.datacamp.com/blog/classification-machine-learning>. [cit. 2023-12-20].
- [9] *Classification Methods in Machine Learning* 2019. Online. In: Medium. Dostupné z: <https://medium.com/@jorgesleoneel/classification-methods-in-machine-learning-58ce63173db8>. [cit. 2023-12-20].
- [10] MAALOUF, Maher, 2011. *Logistic regression in data analysis: An overview*. Online. International Journal of Data Analysis Techniques and Strategies. 3, s. 281-299. DOI: <https://doi.org/10.1504/IJDATS.2011.041335>. [cit. 2023-12-21].
- [11] SPERANDEI, Sandro, 2014. *Understanding logistic regression analysis*. Online. In: Biochem Med (Zagreb). S. 12–18. DOI: <https://doi.org/10.11613/BM.2014.003>. [cit. 2023-12-21].
- [12] JIJO, Abraham. *A Beginner's Guide to K Nearest Neighbor(KNN) Algorithm With Code*. Online. In: Medium. Dostupné z: <https://medium.com/analytics-vidhya/a-beginners-guide-to-k-nearest-neighbor-knn-algorithm-with-code-5015ce8b227e>. [cit. 2023-12-21]
- [13] *K-Nearest Neighbor(KNN) Algorithm*. Online. Geeks for geeks. Dostupné z: <https://www.geeksforgeeks.org/k-nearest-neighbours/>. [cit. 2023-12-21].
- [14] *What is random forest?* Online. IBM. Dostupné z: <https://www.ibm.com/topics/random-forest>. [cit. 2023-12-21].
- [15] MBAABU, Onesmus. *Introduction to Random Forest in Machine Learning*. Online. Section. Dostupné z: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/>. [cit. 2023-12-21].
- [16] *Evaluating a Classification Model*. Online. Ritchie Ng. Dostupné z: <https://www.ritchieng.com/machine-learning-evaluate-classification-model/>. [cit. 2024-02-04].
- [17] *Accuracy vs. precision vs. recall in machine learning: what's the difference?* Online. Evidently AI. Dostupné z: <https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall>. [cit. 2024-02-04].
- [18] KUNDU, Rohit. *F1 Score in Machine Learning: Intro & Calculation*. Online. V7. Dostupné z: <https://www.v7labs.com/blog/f1-score-guide>. [cit. 2024-02-04].

- [19] BROWNLEE PHD, Jason, 2020. *How to Choose a Feature Selection Method For Machine Learning* Online. Machine Learning Mystery. <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>. [cit. 2023-10-23].
- [20] JIE, Cai; JIAWEI, Luo; SHULIN, Wang a SHENG, Yang, 2018. *Feature selection in machine learning: A new perspective*. Online. Neurocomputing. 300, s. 70-79. DOI: <https://doi.org/10.1016/j.neucom.2017.11.077>. [cit. 2023-12-19].
- [21] BOMMERT, Andrea; SUN, Xudong; BISCHL, Bernd; RAHNENFÜHRER, Jörg a LANG, Michel, 2020. *Benchmark for filter methods for feature selection in high-dimensional classification data*. Online. Computational Statistics & Data Analysis. 143. S. 106839. ISSN 0167-9473. DOI: <https://doi.org/10.1016/j.csda.2019.106839>. [cit. 2023-12-18].
- [22] SÁNCHEZ-MAROÑO, Noelia; ALONSO-BETANZOS, Amparo a TOMBILLA-SANROMÁN, María, 2007. *Filter Methods for Feature Selection – A Comparative Study*. Online. Intelligent Data Engineering and Automated Learning - IDEAL 2007. S. 178–187. DOI: https://doi.org/10.1007/978-3-540-77226-2_19. [cit. 2023-12-19].
- [23] STEWART, Ken. *Pearson's correlation coefficient*. Online. In: Britannica. Dostupné z: <https://www.britannica.com/topic/Pearsons-correlation-coefficient>. [cit. 2023-12-28].
- [24] ZHOU, Hongfang; WANG, Xiqian a ZHU, Rourou, 2022. *Feature selection based on mutual information with correlation coefficient*. Online. Applied Intelligence. 52, s. 5457–5474. Dostupné z: <https://link.springer.com/article/10.1007/s10489-021-02524-x>. [cit. 2023-12-19].
- [25] *Chi-Square Test for Feature Selection – Mathematical Explanation*. Online. In: Geeks for geeks. Dostupné z: <https://www.geeksforgeeks.org/chi-square-test-for-feature-selection-mathematical-explanation/>. [cit. 2023-12-19].
- [26] TZINI, Elli. *Feature Selection: Wrapper Methods: 5 Wrapper-based methods to choose relevant features*. Online. In: Medium. Dostupné z: <https://medium.com/analytics-vidhya/feature-selection-85539d6a2a88>. [cit. 2023-12-20].
- [27] KOHAVI, Ron a JOHN, George, 1997. *Wrappers for feature selection*. Online. Artificial Intelligence - AI. 1. Dostupné z:

- https://www.researchgate.net/publication/243768287_Wrappers_for__feature_selection. [cit. 2023-12-20].
- [28] *Introduction to Dimensionality Reduction*. Online. Geeks for geeks. Dostupné z: <https://www.geeksforgeeks.org/dimensionality-reduction/>. [cit. 2024-02-10].
- [29] *A Step-by-Step Explanation of Principal Component Analysis (PCA)*. Online. In: BuiltIn. Dostupné z: <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>. [cit. 2023-12-20].
- [30] LEE, Pyeong-Yeon; KWON, Sanguk; KANG, Deokhun; KIM, Jonghoon a CHO, Inho, 2022. *Principle component analysis-based optimized feature extraction merged with nonlinear regression model for improved state-of-health prediction*. Online. Journal of Energy Storage. 48, s. 104026. DOI: <https://doi.org/10.1016/j.est.2022.104026>. [cit. 2023-12-20].
- [31] BALAN, Srinivasan *Metaheuristics in Optimization: Algorithmic Perspective* Online. Informs. Copyright 2023 INFORMS. <https://www.informs.org/Publications/OR-MS-Tomorrow/Metaheuristics-in-Optimization-Algorithmic-Perspective>. [cit. 2023-10-24].
- [32] YANG, Xin-She, 2014. *Nature-Inspired Optimization Algorithms* Elsevier. ISBN 9780124167438.
- [33] MISHRA, Sushruta; MALLICK, Pradeep Kumar; CHAE, Gyoo-Soo; TRIPATHY, Hrudaya Kumar a SANGAIAH, Arun Kumar, 2022. *Cognitive Big Data Intelligence with a Metaheuristic Approach* Academic Press. ISBN 9780323851176.
- [34] CHENG, Ling; ABU-MAHFOUZ, Adnan M; HOFSAJER, Ivan W. a ADETUNJI, Kayode E., 2021. *A Review of Metaheuristic Techniques for Optimal Integration of Electrical Units in Distribution Networks*. Online. IEEE Access. 9, s. 5046-5068. DOI: <https://doi.org/10.1109/ACCESS.2020.3048438>. [cit. 2024-02-07].
- [35] YANG, Xin-She; CUI, Zhihua; HOSSEIN GANDOMI, Amir a KARAMANOGLU, Mehmet, 2013. *Swarm intelligence and bio-inspired computation*. Elsevier. ISBN 987-0-12-405163-8.
- [36] YANG, Xin-She, 2010. *Nature-Inspired Metaheuristic Algorithms*. 2nd Edition. Luniver Press. ISBN 9781905986286.

- [37] PENTAPALLI, Venkata Vijaya Geeta a VARMA P, Ravi Kiran, 2016. *Cuckoo Search Optimization and its Applications: A Review* Online. International Journal of Advanced Research in Computer and Communication Engineering. 5, s. 7. ISSN 2278-1021. DOI: <https://doi.org/10.17148/IJARCCCE.2016.511119>. [cit. 2023-10-24].
- [38] YANG, Xin-She a DEB, Suash, 2014. *Cuckoo search: recent advances and applications* Online. Neural Computing and Applications. 24, s. 169–174. DOI: <https://doi.org/10.1007/s00521-013-1367-1>. [cit. 2023-10-28].
- [39] YANG, Xin-She, 2011. *Metaheuristic Optimization* Online. Scholarpedia. Dostupné z: http://www.scholarpedia.org/article/Metaheuristic_Optimization. [cit. 2023-11-10].
- [40] YANG, Xin-She a DEB, Suash, 2019. *Cuckoo Search via Levy Flights* Online. Proc. of World Congress on Nature & Biologically Inspired Computing. s. 210-214. Dostupné z: https://www.researchgate.net/publication/45904981_Cuckoo_Search_via_Levy_Flights. [cit. 2023-11-10].
- [41] MOHAMAD, Azizah Binti; ZAIN, Azlan Mohd a BAZIN, Nor Erne Nazira, 2014. *Cuckoo Search Algorithm for Optimization Problems—A Literature Review and its Applications* Online. Applied Artificial Intelligence. Issue 5, 28, s. 419-448. DOI : <https://doi.org/10.1080/08839514.2014.904599>. [cit. 2023-11-10].
- [42] HEIDARI, Ali Asghar; MIRJALILI, Seyedali; FARIS, Hossam; ALJARAH, Ibrahim; MAFARJA, Majdi et al., 2019. *Harris hawks optimization: Algorithm and applications* Online. Future Generation Computer Systems. 97, s. 849-872. ISSN 0167-739X. DOI: <https://doi.org/10.1016/j.future.2019.02.028>. [cit. 2023-10-28].
- [43] GEZICI, Harun a LIVATYALI, Haydar, 2022. *Chaotic Harris hawks optimization algorithm* Online. Journal of Computational Design and Engineering. 9, issue 1, 216-245. DOI: <https://doi.org/10.1093/jcde/qwab082>. [cit. 2023-11-10].
- [44] ALABOOL, Hamzeh Mohammad; ALARABIAT, Deemah; ABUALIGAH, Laith a HEIDARI, Ali Asghar, 2021. *Harris hawks optimization: a comprehensive review of recent variants and applications* Online. Neural Computing and Applications. 33, s. 8939–8980. DOI: <https://doi.org/10.1007/s00521-021-05720-5>. [cit. 2023-11-10].

- [45] SHEHAB, Mohammad; MASHAL, Ibrahim; MOMANI, Zaid; SHAMBOUR, Mohd Khaled Yousef; AL-BADAREEN, Anas et al., 2022. *Harris Hawks Optimization Algorithm: Variants and Applications* Online. Archives of Computational Methods in Engineering. 29, s. 5579–5603. DOI: <https://doi.org/10.1007/s11831-022-09780-1>. [cit. 2023-11-10].
- [46] TRIPATHY, B. K.; MADDIKUNTA, Praveen Kumar Reddy; PHAM, Quoc-Viet; GADEKALLU, Thippa Reddy; DEV, Kapal et al., 2022. *Harris Hawk Optimization: A Survey on Variants and Applications* Online. Computational Intelligence and Neuroscience. 20. DOI: <https://doi.org/10.1155/2022/2218594>. [cit. 2023-11-10].
- [47] RANA, Nadim; LATIFF, Muhammad Shafie Abd; ABDULHAMID, Šáfií Muhammad a CHIROMA, Haruna, 2020. *Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments*. Online. Neural Comput & Applic. 32, s.16245–16277. DOI: <https://doi.org/10.1007/s00521-020-04849-z>. [cit. 2023-11-03].
- [48] BRODZICKI, Andrzej; PIEKARSKI, Michał a JAWOREK-KORJAKOWSKA, Joanna, 2021. *The Whale Optimization Algorithm Approach for Deep Neural Networks* Online. Sensors 2021, 21(23), 8003. DOI: <https://doi.org/10.3390/s21238003>. [cit. 2023-11-04].
- [49] JUNBO, Ling; MA, Guohua; LIAN, Ling; ZHU, Ting; WU, Xincan et al. *Parrot optimizer: Algorithm and applications to medical problems*. Online. Computers in Biology and Medicine. 172, s. 108064. DOI: <https://doi.org/10.1016/j.compbimed.2024.108064>. [cit. 2024-04-06].
- [50] JIA, Yunjian; WU, Yanfei; LIANG, Liang; WEI, Yaxing a WANG, Shankun, 2023. *A Flower Pollination Optimization Algorithm Based on Cosine Cross-Generation Differential Evolution*. Online. Sensors. 23. ISSN 1424-8220. DOI: doi.org/10.3390/s23020606. [cit. 2024-04-07].
- [51] XIN-SHE, Yang. *Flower Pollination Algorithm for Global Optimization*. Online. Unconventional Computation and Natural Computation. Roč. 2012, 7445, s. 240-249. DOI: https://doi.org/10.1007/978-3-642-32894-7_27. [cit. 2024-04-07].
- [52] KARABOGA, Dervis; GORKEMLI, Beyza; OZTURK, Celal a KARABOGA, Nurhan, 2014. *A comprehensive survey: artificial bee colony (ABC) algorithm and applications*. Online. Artificial Intelligence Review. 42, 21-57. DOI: <https://doi.org/10.1007/s10462-012-9328-0>. [cit. 2023-11-18].

- [53] ZHOU, Xinyu; LU, Jiaxin; HUANG, Junhong; ZHONG, Maosheng a WANG, Mingwen, 2021. *Enhancing artificial bee colony algorithm with multi-elite guidance*. Online. Information Sciences. 543, 242-258. DOI: <https://doi.org/10.1016/j.ins.2020.07.037>. [cit. 2023-11-18].
- [54] PHAM, D. T.; GHANBARZADEH, Afshin; KOÇ, Ebubekir; OTRI, Sameh; RAHIM, Sahra et al., 2005. *Bee Algorithm A Novel Approach to Function Optimisation*. Online. Manufacturing Engineering Centre, Cardiff University, UK. S. 1-57. Dostupné z: https://www.researchgate.net/publication/260985621_The_Bees_Algorithm_Technical_Note. [cit. 2023-11-18].
- [55] STÜTZLE, Thomas a DORIGO, Marco, 2018. *Ant Colony Optimization: Overview and Recent Advances*. Online. Handbook of Metaheuristics. 272. Springer, Cham, s. 311-351. ISBN 978-3-319-91086-4. Dostupné z: https://link.springer.com/chapter/10.1007/978-3-319-91086-4_10. [cit. 2023-11-20].
- [56] SHTOVBA, Serhiy, 2005. *Ant Algorithms: Theory and Applications*. Online. Programming and Computer Software. Pleiades Publishing, s. 167-178. ISBN 1608-3261. Dostupné z: https://www.researchgate.net/publication/220203867_Ant_Algorithms_Theory_and_Applications. [cit. 2023-11-20].
- [57] DENG, Wu; XU, Junjie a ZHAO, Huimin, 2019. *An Improved Ant Colony Optimization Algorithm Based on Hybrid Strategies for Scheduling Problem*. Online. IEEE Access. 7, 20281-20292. DOI: <https://doi.org/10.1109/ACCESS.2019.2897580>. [cit. 2023-11-20].
- [58] KATIYAR, Sapna; ANSARI, Abdul Quaiyum a NASIRUDDIN, Ibraheem. *Ant Colony Optimization: A Tutorial Review*. Online. https://www.researchgate.net/publication/281432201_Ant_Colony_Optimization_A_Tutorial_Review. [cit. 2023-11-20].
- [59] BARBOSA, Helio, 2013. *Ant Colony Optimization - Techniques and Applications*. Online. IntechOpen. ISBN 978-953-51-5717-5. DOI: <https://doi.org/10.5772/3423>. [cit. 2023-11-21].
- [60] FARAHLINA JOHARI, Nur; MOHD ZAIN, Azlan; MUSTAFFA, Norfa a UDIN, Amirmudin, 2013. *Firefly Algorithm for Optimization Problem*. Online. Applied Mechanics and Materials. 421, s. 512-517. DOI: <https://doi.org/10.4028/www.scientific.net/AMM.421.512>. [cit. 2023-12-22].

- [61] KUMAR, Vijay a KUMAR, Dinesh, 2021. *A Systematic Review on Firefly Algorithm: Past, Present, and Future*. Online. Archives of Computational Methods in Engineering. 28, s. 3269–3291. DOI: <https://doi.org/10.1007/s11831-020-09498-y>. [cit. 2023-12-22].
- [62] KAVITA, Shelke a S. K., Shinde, 2023. *Metaheuristic Evolutionary Algorithms: Types, Applications, Future Directions, and Challenges*. Online. 3rd International Conference on Intelligent Technologies. S. 1-6. DOI: <https://doi.org/10.1109/CONIT59222.2023.10205592>. [cit. 2023-12-22].
- [63] *Genetic Algorithm Based on Natural Selection Theory for Optimization Problems*, 2020. Online. MDPI. Dostupné z: https://www.mdpi.com/symmetry/symmetry-12-01758/article_deploy/html/images/symmetry-12-01758-g001-550.jpg. [cit. 2023-12-22].
- [64] ALHIJAWI, Bushra a AWAJAN, Arafat, 2023. *Genetic algorithms: theory, genetic operators, solutions, and applications*. Online. Evolutionary Intelligence. DOI: <https://doi.org/10.1007/s12065-023-00822-6>. [cit. 2023-12-22].
- [65] OPARA, Karol R. a ARABAS, Jarosław, 2019. *Differential Evolution: A survey of theoretical analyses*. Online. Swarm and Evolutionary Computation. 44, s. 546-558. DOI: <https://doi.org/10.1016/j.swevo.2018.06.010>. [cit. 2024-02-11].
- [66] AHMAD, Mohamad Faiz; MAT ISA, Nor Ashidi; LIM, Wei Hong a ANG, Koon Meng, 2022. *Differential evolution: A recent review based on state-of-the-art works*. Online. Alexandria Engineering Journal. 5, 61, s. 3831-3872. DOI: <https://doi.org/10.1016/j.aej.2021.09.013>. [cit. 2024-02-11].
- [67] SUCHOMEL, Ondřej a HYHLÍK, Tomáš. *Optimalizační metody v CFD – diferenciální evoluce*. Online. Dostupné z: <http://www.suchomel.org/files/Optimalizace-v-CFD.pdf>. [cit. 2024-02-12].
- [68] BISWAS, Anupam; MISHRA, K. K.; TIWARI, Shailesh a MISRA, A. K., 2013. *Physics-Inspired Optimization Algorithms: A Survey*. Online. Journal of Optimization. DOI: <https://doi.org/10.1155/2013/438152>. [cit. 2023-12-22].
- [69] MONGEAU, Marcel; CHAIMATANAN, Supatcha a DELAHAYE, Daniel, 2018. *Simulated Annealing: From Basics to Applications*. Online. Handbook of Metaheuristics. 272, s. 1-35. Dostupné z: https://link.springer.com/chapter/10.1007/978-3-319-91086-4_1. [cit. 2023-12-22].

- [70] Hunt, J. (2023). *Python Type Hints*. In: *Advanced Guide to Python 3 Programming. Undergraduate Topics in Computer Science*. Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-40336-1_2. [cit. 2023-09-05].
- [71] JANOSI, Andras; STEINBRUNN, William; PFISTERER, Matthias a DETRANO, Robert, 1988. *Heart Disease* Online. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C52P4X>. [cit. 2023-12-01].
- [72] *Ucimplrepo*. Online. Dostupné z: <https://github.com/uci-ml-repo/ucimplrepo>. [cit. 2023-12-27].
- [73] *Heart failure clinical records* 2020. Online. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5Z89R>. [cit. 2024-02-13].
- [74] PRAKASH DOPPALA, Bhanu a BHATTACHARYYA, Debnath, 2021. *Cardiovascular_Disease_Dataset*. Online. Mendeley Data. DOI: <https://doi.org/10.17632/dzz48mvjht.1>. [cit. 2024-02-17].
- [75] PEDREGOSA, Fabian; VAROQUAUX, Gael; GRAMFORT, Alexandre; VINCENT, Michel; THIRION, Bertrand et al., 2012. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research. 12, s. 2826-2830.
- [76] PATRO, Siboprasad; PADHY, Neelamadhab a DEO SAH, Rahul, 2022. *An improved ensemble learning approach for the prediction of cardiovascular disease using majority voting prediction*. Online. Modeling, identification and control. 41, s. 68-86. ISSN 1746-6172. DOI: <https://doi.org/10.1504/IJMIC.2022.127098>. [cit. 2024-05-21].
- [77] G, Ambrish; GANESH, Bharathi; GANESH, Anitha; SRINIVAS, Chetana a MENSINKAL, Kiran, 2022. *Logistic regression technique for prediction of cardiovascular disease*. Online. Global Transitions Proceedings. 3, s. 127-130. DOI: <https://doi.org/10.1016/j.gltp.2022.04.008>. [cit. 2024-05-21].
- [78] ZHENYA, Qi a ZHANG, Zuoru, 2021. *A hybrid cost-sensitive ensemble for heart disease prediction*. Online. BMC Medical Informatics and Decision Making. 21. DOI: <https://doi.org/10.1186/s12911-021-01436-7>. [cit. 2024-05-21].
- [79] HERA, Sabeena Yasmin; AMJAD, Mohammad a SABA, Md Khalaf, 2022. *Improving heart disease prediction using multi-tier ensemble model*. Online. Network Modeling Analysis in Health Informatics and Bioinformatics. 11. DOI: <https://doi.org/10.1007/s13721-022-00381-3>. [cit. 2024-05-21].

- [80] CHICCO, Davide a JURMAN, Giuseppe, 2020. *Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone*. Online. BMC Medical Informatics and Decision Making. DOI: <https://doi.org/10.1186/s12911-020-1023-5>. [cit. 2024-05-21].
- [81] ISHAQ, Abid; SADIQ, Saima; UMER, Muhammad a ULLAH, Saleem, 2021. *Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques*. Online. IEEE Access. S. 39707 - 39716. DOI: <https://doi.org/10.1109/ACCESS.2021.3064084>. [cit. 2024-05-21].
- [82] NISHAT, Mirza Muntasir; FAISAL, Fahim; ABRAR, Al-Monsur; AR-RAFI, Mohammad a RATUL, Ishrak Jahan, 2022. *A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset*. Online. Hindawi. DOI: <https://doi.org/10.1155/2022/3649406>. [cit. 2024-05-21].
- [83] KARAKUŞ, Mücella Özbay a ER, Orhan, 2022. *A comparative study on prediction of survival event of heart failure patients using machine learning algorithms*. Online. Neural Computing and Applications. 34, s. 13895–13908. DOI: <https://doi.org/10.1007/s00521-022-07201-9>. [cit. 2024-05-21].
- [84] DOPPALA, BP; BHATTACHARYYA, D; JANARTHANAN, M a BALIK, N, 2022. *A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques*. Online. Journal of Healthcare Engineering. DOI: <https://doi.org/10.1155/2022/2585235>. [cit. 2024-05-21].
- [85] ANU, Radha.P a DAVID, Vasantha Kalyani, 2021. *SUPER LEARNER MODEL IN PREDICTION OF HEART ATTACK BASED ON CARDIAC BIOMARKERS*. Online. Indian Journal of Computer Science and Engineering. 12, s. 1702-1712. DOI: <https://doi.org/10.21817/indjce/2021/v12i6/211206076>. [cit. 2024-05-21].

Seznam symbolů a zkratek

ABC	Algoritmus umělého včelstva - artificial bee colony
ACO	Algoritmus kolonie mravenců - ant colony optimization
CS	Algoritmus podle kukaček - cuckoo search
DE	Diferenciální evoluce - differential evolution
FA	Algoritmus světlušky - firefly algorithm
FN	Falešně negativní - false negative
FP	Falešně pozitivní - false positive
FPA	Algoritmus opylování květin - flower pollination algorithm
GA	Genetický algoritmus - genetic algorithm
HHO	Algoritmus podle jestřábů Harisových - Harris hawks optimization
K-NN	Algoritmus nejbližších sousedů - K-Nearest neighbors algorithm
LR	Logistická regrese - logistic regression
ML	Strojové učení - machine learning
RF	Náhodná doménová struktura - random forest
SA	Simulované žíhání - simulated annealing
TN	Skutečně negativní - true negative
TP	Skutečně pozitivní - true positive
WOA	Algoritmus podle velryb - whale optimization algorithm

Seznam příloh

A Klevelandský dataset	72
B Dataset srdečního selhání	73
C Indický dataset	74
D Pseudokódy použitých datasetů	75
E Souhrnné výsledky pro vyzkoušené nastavení parametrů	78

A Klevelandský dataset

Níže je uvedena tabulka příznaků tohoto datasetu s odpovídajícím rozsahem hodnot před úpravou.

Tab. A.1: Tabulka příznaků s rozsahem hodnot pro Klevelandský dataset

Příznak	Datový typ	Vysvětlení	Rozsah
age	číselná	věk pacienta v letech	29 - 77
sex	kategorická	pohlaví pacienta	0 - žena 1 - muž
cp	kategorická	typ bolesti na hrudi	1 - typická angina pectoris 2 - atypická angina pectoris 3 - jiná bolest 4 - asymptomatický
trestbps	číselná	klidový systolický krevní tlak [mmHg]	94 - 200
chol	číselná	cholesterol v séru [mg/dl]	126 - 564
fbs	kategorická	glykémie na lačno [mg/dl]	0 - nižší než 120 1 - vyšší než 120
restecg	kategorická	klidové EKG vyšetření	0 - fyziologické 1 - abnormalita ST úseku
thalach	číselná	max. tepová frekvence [tep/min]	71-202
exang	kategorická	angina pectoris vyvolaná fyzickou aktivitou	0 - ne 1 - ano
oldpeak	číselná	deprese ST úseku [mm]	0 - 6,2
slope	kategorická	sklon ST úseku při maximální zátěži	0 - vzestupný 1 - plochý 2 - sestupný
ca	číselná	počet hlavních koronárních cév postižených stenózou	0 - 3
thal	kategorická	talasémie - dědičné onemocnění krve	0 - normální 1 - opravený defekt 2 - reverzibilní defekt
num	kategorická	závažnost srdečního onemocnění na základě příznaků	0 - bez onemocnění 1 - velmi nízká závažnost 2 - nízká závažnost 3 - střední závažnost 4 - vysoká závažnost

B Dataset srdečního selhání

Níže je uvedena tabulka příznaků tohoto datasetu s odpovídajícím rozsahem hodnot před úpravou.

Tab. B.1: Tabulka příznaků s rozsahem hodnot pro dataset srdečního selhání

Příznak	Datový typ	Vysvětlení	Rozsah
age	číslná	věk pacienta v letech	40 - 95
anemia	kategorická	snížení počtu červených krvinek	0 - ne 1 - ano
creatinine phosphokinase	číslná	hodnota kreatinfosfokinázy v krvi [mcg/l]	23 - 7861
diabetes	kategorická	pacient má diabetes	0 - ne 1 - ano
ejection fraction	číslná	objem krve vypuzené srdcem při kontrakci [%]	14 - 80
high blood pressure	kategorická	pacient má hypertenzi (vysoký krevní tlak)	0 - ne 1 - ano
platelets	číslná	počet krevních destičky v krvi [počet/ μ l]	25,100 - 850,000
sex	kategorická	pohlaví pacienta	0 - muž 1 - žena
serum creatinine	číslná	obsah kreatininu v séru [mg/dl]	0,5 - 9,4
serum sodium	číslná	obsah sodíku v séru [mmol/l]	113 - 148
smoking	kategorická	pacient je kuřák	0 - ne 1 - ano
time	číslná	sledovací období [dny]	4 - 285
death event	kategorická	pacient umřel během sledovacího období	0 - ne 1 - ano

C Indický dataset

Níže je uvedena tabulka příznaků tohoto datasetu s rozsahem hodnot před úpravou.

Tab. C.1: Tabulka příznaků s rozsahem hodnot pro Indický dataset

Příznak	Datový typ	Vysvětlení	Rozsah
age	číselná	věk pacienta v letech	20 - 80
gender	kategorická	pohlaví pacienta	0 - žena 1 - muž
chestpain	kategorická	typ bolesti na hrudi	0 - typická angina pectoris 1 - atypická angina pectoris 2 - jiná bolest 3 - asymptomatický
restingBP	číselný	klidový systolický krevní tlak [mmHg]	94 - 200
serumcholesterol	číselná	cholesterol v krvi [mg/dl]	0 - 602
fasting blood sugar	kategorická	glykémie na lačno [mg/dl]	0 - nižší než 120 1 - vyšší než 120
restingelectro	kategorická	klidové EKG vyšetření	0 - fyziologické 1 - abnormalita ST úseku 2 - hypertrofie levé komory
maxheartrate	číselná	maximální tepová frekvence [tep/min]	71 - 202
exerciseangia	kategorická	angina pectoris vyvolaná fyzickou aktivitou	0 - ne 1 - ano
oldpeak	číselná	deprese ST úseku vyvolaná cvičením [mm]	0 - 6,2
slope *	kategorická	sklon ST úseku při maximální zátěži	0 - fyziologický 1 - vzestupný 2 - plochý 3 - sestupný
noofmajorvessels	číselná	počet hlavních koronárních cév postižených stenózou	0 - 3
target	kategorická	výskyt srdečního onemocnění	0 - zdravý 1 - nemocný

*V původní dokumentaci k datasetu se nacházejí pouze tři kategorie, při práci s daty bylo zjištěno, že se zde nachází čtyři kategorie (navíc fyziologický úsek ST).

D Pseudokódy použitých datasetů

Algoritmus kukačky 3.1.1

```
Inicializace parametrů
Inicializace počáteční populace s určením fitness hodnoty
for  $i = 1$  : max iterace do
    Aktualizace populace pomocí Levyho letu (3.1)
    Fitness hodnoty pro novou populaci
    for  $j = 1$  : pocet jedincu do
        if  $r < pa$  then
            Opuštění hnízda a náhodné postavení nového hnízda
        else if  $r \geq pa$  then
            Ponechání jedince  $j$ 
        end if
    end for
end for
Návrat nejlepšího jedince (řešení)
end
```

Algoritmus jestřábů Harrisových (3.1.2)

```
Inicializace parametrů
Inicializace počáteční populace s určením fitness hodnoty
for  $i = 1$  : max iterace do
    for  $j = 1$  : pocet jedincu do
        if  $E \geq 1$  then
            Průzkumná fáze
            Aktualizace jedince  $j$  dle rovnice (3.3)
        else if  $E < 1$  then
            else if  $r \geq 0,5$  and  $E \geq 0,5$  then
                Měkké obléhání
                Aktualizace jedince  $j$  dle rovnice (3.6)
            else if  $r \geq 0,5$  and  $E < 0,5$  then
                Tvrdé obléhání
                Aktualizace jedince  $j$  dle rovnice (3.9)
            else if  $r < 0,5$  and  $E \geq 0,5$  then
                Ponory
                Aktualizace jedince  $j$  dle rovnice (3.8)
            else if  $r < 0,5$  and  $E < 0,5$  then
                Měkké obléhání s nálety
                Aktualizace jedince  $j$  dle rovnice (3.7)
            end if
        end for
    end for
Návrat nejlepšího jedince (řešení)
end
```

Algoritmus velryby 3.1.3

```
Inicializace parametrů
Inicializace počáteční populace s určením fitness hodnoty
for  $i = 1$  : max iterace do
  for  $j = 1$  : pocet jedincu do
    if  $p < 0.5$  then
      if  $A \geq 1$  then
        Fáze průzkumu
        Aktualizace jedince  $j$  dle rovnice (3.10, 3.11)
      else if  $A < 1$  then
        Fáze obklopování kořisti
        Aktualizace jedince  $j$  dle rovnice (3.14, 3.15)
      end if
    else if  $p \geq 0.5$  then
      Fáze útoku bublinovou sítí
      Aktualizace jedince  $j$  dle rovnice (3.16)
    end if
  end for
end for
Návrat nejlepšího jedince (řešení)
end
```

Algoritmus papouška 3.1.4

```
Inicializace parametrů
Inicializace počáteční populace s určením fitness hodnoty
for  $i = 1$  : max iterace do
  for  $j = 1$  : pocet jedincu do
     $St \leftarrow \text{randi}([1, 4])$ 
    if  $St == 1$  then
      Stav 1: krmení
      Aktualizace jedince  $j$  dle rovnice (3.18)
    else if  $St == 2$  then
      Stav 2: sezení na chovateli
      Aktualizace jedince  $j$  dle rovnice (3.20)
    else if  $St == 3$  then
      Stav 3: komunikace v hejnu
      Aktualizace jedince  $j$  dle rovnice (3.21)
    else if  $St == 4$  then
      Stav 4: strach z nepřítele
      Aktualizace jedince  $j$  dle rovnice (3.22)
    end if
  end for
end for
Návrat nejlepšího jedince (řešení)
end
```

Algoritmus opylování květin 3.1.5

```
Inicializace parametrů
Inicializace počáteční populace s určení fitness hodnoty
for  $i = 1$  : max iterace do
  for  $j = 1$  : pocet jedincu do
    if  $r < p$  then
      Globální prohledávání
      Aktualizace jedince  $j$  dle rovnice (3.23)
    else if  $r \geq p$  then
      Lokální prohledávání
      Aktualizace jedince  $j$  dle rovnice (??)
    end if
  end for
end for
Návrat nejlepšího jedince (řešení)
end
```

E Souhrnné výsledky pro vyzkoušené nastavení parametrů

Tab. E.1: Výsledky PO pro Klevelandský dataset

KLEVELANDSKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,8636	0,8181	0,8667	0,8947	0,8181	0,7272	0,8000	0,8421	0,7272
Příznaky	sex, chol, restecg, thalach, oldpeak, ca, cp_2, slope_1, slope_0, thal_1, thal_0, thal_2								
Alfa = 1	0,8780	0,7907	0,8500	0,8947	0,7727	0,7727	0,8333	0,8684	0,7727
Příznaky	age, sex, restecg, oldpeak, ca, cp_1, cp_3, cp_2, slope_2, slope_1, thal_1								
Alfa = 1,5	0,8837	0,7441	0,8167	0,8684	0,7272	0,7907	0,8500	0,8947	0,7727
Příznaky	age, sex, fbs, restecg, exang, ca, cp_2, slope_2, slope_1								
Parametry	30 jedinců								
Alfa = 0,5	0,9000	0,7441	0,8166	0,8684	0,7272	0,7179	0,8166	0,9220	0,6363
Příznaky	sex, restecg, thalach, exang, oldpeak, ca, cp_4, cp2, slope_2, slope_1, thal_1, thal_0								
Alfa = 1	0,8636	0,8095	0,8666	0,9211	0,7727	0,7442	0,8167	0,8684	0,7272
Příznaky	sex, trestbps, chol, restecg, thalach, exang, oldpeak, ca, cp_1, cp_4, cp2, slope_2, slope_1, slope_0, thal_2								
Alfa = 1,5	0,9000	0,8000	0,8500	0,8684	0,8181	0,7499	0,8333	0,9211	0,6818
Příznaky	sex, restecg, thalach, exang, oldpeak, ca, cp2, slope_1, thal_1, thal_2								
Parametry	60 jedinců								
Alfa = 0,5	0,8837	0,8372	0,8833	0,9211	0,8181	0,8000	0,8500	0,8684	0,8181
Příznaky	sex, trestbps, chol, fbs, restecg, thalach, oldpeak, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1								
Alfa = 1	0,9048	0,8181	0,8666	0,8947	0,8181	0,7907	0,8500	0,8947	0,7727
Příznaky	sex, trestbps, chol, fbs, restecg, thalach, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, thal_2								
Alfa = 1,5	0,8780	0,7826	0,8333	0,8421	0,8181	0,7442	0,8166	0,8684	0,7272
Příznaky	age, sex, trestbps, fbs, restecg, ca, cp_2, slope_2, slope_1, slope_0, thal_1								
Parametry	100 jedinců								
Alfa = 0,5	0,8837	0,8444	0,8833	0,8947	0,8636	0,7907	0,8500	0,8947	0,7727
Příznaky	sex, chol, restecg, thalach, ca, cp_2, slope_2, slope_0, thal_1								
Alfa = 1	0,9048	0,8000	0,8500	0,8684	0,8181	0,8293	0,8833	0,9474	0,7727
Příznaky	age, sex, chol, fbs, restecg, oldpeak, ca, cp_2, slope_2, slope_0, thal_1								
Alfa = 1,5	0,8780	0,8000	0,8500	0,8684	0,8181	0,7727	0,8333	0,8684	0,7727
Příznaky	age, sex, fbs, restecg, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1								

Tab. E.2: Výsledky PO pro Dataset srdečního selhání

DATASET SRDEČNÍHO SELHÁNÍ									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,8000	0,6250	0,7895	0,9722	0,4762	0,6857	0,8070	0,9444	0,5714
Příznaky	ejection_fraction , platenes , trestbps , serum_sodium , time								
Alfa = 1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7428	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , platelets , serum_sodium , smoking , time								
Alfa = 1,5	0,8108	0,7059	0,8246	0,9722	0,5714	0,6660	0,7895	0,9167	0,5714
Příznaky	ejection_fraction , high_blood_pressure , serum_sodium , time								
Parametry	30 jedinců								
Alfa = 0,5	0,8205	0,7058	0,8246	0,9722	0,5714	0,6286	0,7719	0,9166	0,5238
Příznaky	anemia , ejection_fraction , high_blood_pressure , platelets , serum_sodium , time								
Alfa = 1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , platelets , serum_sodium , smoking , time								
Alfa = 1,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , platelets , serum_sodium , smoking , time								
Parametry	60 jedinců								
Alfa = 0,5	0,8108	0,6286	0,7719	0,9166	0,5238	0,6666	0,7895	0,9166	0,5714
Příznaky	age , ejection_fraction , high_blood_pressure , serum_sodium , smoking , time								
Alfa = 1	0,7805	0,6061	0,7719	0,9444	0,4762	0,7805	0,8421	0,8888	0,7619
Příznaky	ejection_fraction , serum_creatinine , time								
Alfa = 1,5	0,8333	0,6666	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , platelets , serum_sodium , smoking , time								
Parametry	100 jedinců								
Alfa = 0,5	0,8108	0,7059	0,8246	0,9722	0,5714	0,6666	0,7895	0,9166	0,5714
Příznaky	ejection_fraction , high_blood_pressure , serum_sodium , time								
Alfa = 1	0,7805	0,6060	0,7719	0,9444	0,4762	0,7805	0,8421	0,8888	0,7619
Příznaky	ejection_fraction , serum_creatinine , time								
Alfa = 1,5	0,8205	0,6471	0,7895	0,9444	0,5238	0,3871	0,6666	0,8888	0,2857
Příznaky	age , anemia , ejection_fraction , high_blood_pressure , platelets , serum_sodium , sex , time								

Tab. E.3: Výsledky PO pro Indický dataset

INDICKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	1,0000	0,9741	0,9700	0,9759	0,9658	0,9565	0,9500	0,9638	0,9402
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingelectro_1, restingelectro_2								
Alfa = 1	0,9957	0,9741	0,9700	0,9760	0,9658	0,9576	0,9500	0,9277	0,9658
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_0, chestpain_1, chestpain_3, slope_1, slope_2, slope_0, restingelectro_0, restingelectro_2								
Alfa = 1,5	1,0000	0,9741	0,9700	0,9759	0,9658	0,9620	0,9550	0,9277	0,9743
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, noofmajorvessels, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingelectro_0, restingelectro_2								
Parametry	30 jedinců								
Alfa = 0,5	0,9957	0,9741	0,9700	0,9759	0,9658	0,9617	0,9550	0,9398	0,9658
Příznaky	gender, restingBP, fastingbloodsugar, maxheartrate, chestpain_2, chestpain_1, chestpain_3, slope_3, slope_2, slope_0, restingelectro_1, restingelectro_2								
Alfa = 1	1,0000	0,9699	0,9650	0,9639	0,9658	0,9623	0,9550	0,9157	0,9829
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_1, slope_3, slope_2, slope_0, restingelectro_0, restingelectro_2								
Alfa = 1,5	1,0000	0,9783	0,9750	0,9879	0,9658	0,9699	0,9650	0,9638	0,9658
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingelectro_2								
Parametry	60 jedinců								
Alfa = 0,5	0,9915	0,9705	0,9650	0,9397	0,9829	0,9699	0,9650	0,9638	0,9658
Příznaky	gender, serumcholesterol, fastingbloodsugar, oldpeak, chestpain_2, chestpain_3, slope_3, slope_1, slope_0, restingelectro_2								
Alfa = 1	0,9957	0,9785	0,9750	0,9759	0,9744	0,9532	0,9450	0,9277	0,9573
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_2, slope_0, restingelectro_1, restingelectro_2								
Alfa = 1,5	0,9914	0,9696	0,9650	0,9759	0,9573	0,9741	0,9700	0,9759	0,9658
Příznaky	age, restingBP, fastingbloodsugar, maxheartrate, chestpain_0, chestpain_1, slope_2, slope_0, restingelectro_2								
Parametry	100 jedinců								
Alfa = 0,5	0,9957	0,9658	0,9600	0,9581	0,9658	0,9828	0,9800	0,9880	0,9744
Příznaky	serumcholesterol, fastingbloodsugar, noofmajorvessels, chestpain_1, slope_1, slope_2, slope_0, restingelectro_1, restingelectro_2								
Alfa = 1	0,9957	0,9785	0,9750	0,9759	0,9744	0,9661	0,9600	0,9398	0,9743
Příznaky	age, gender, restingBP, serumcholesterol, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, restingelectro_0, restingelectro_2								
Alfa = 1,5	0,9957	0,9785	0,9750	0,9759	0,9744	0,9748	0,9700	0,9398	0,9914
Příznaky	age, gender, restingBP, serumcholesterol, maxheartrate, oldpeak, chestpain_2, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								

Tab. E.4: Výsledky HHO pro Klevelandský dataset

KLEVELANDSKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,7755	0,6122	0,6833	0,6842	0,6818	0,6249	0,7000	0,7105	0,6818
Příznaky	trestbps , chol , restecg , exang , oldpeak , ca , slope_0 , thal_0								
Alfa = 1	0,8095	0,6818	0,7667	0,8158	0,6818	0,7442	0,8167	0,8684	0,7272
Příznaky	sex , trestbps , chol , restecg , thalach , exang , ca , cp_2 , slope_2 , slope_1 , slope_0 , thal_0								
Alfa = 1,5	0,7755	0,6872	0,6188	0,6822	0,6832	0,6249	0,7224	0,7272	0,7042
Příznaky	age , sex , fbs , restecg , exang , ca , cp_2 , slope_2								
Parametry	30 jedinců								
Alfa = 0,5	0,7000	0,7000	0,8000	0,8400	0,6363	0,7000	0,8000	0,8947	0,6363
Příznaky	chol , thal_2								
Alfa = 1	0,8372	0,8182	0,8667	0,8947	0,8182	0,7272	0,8000	0,8421	0,7272
Příznaky	age , sex , trestbps , chol , fbs , restecg , thalach , ca , cp2 , slope_2 , thal_1								
Alfa = 1,5	0,6686	0,7222	0,8400	0,8822	0,6556	0,7222	0,7682	0,8889	0,6400
Příznaky	sex , ca , thal_2								
Parametry	60 jedinců								
Alfa = 0,5	0,7000	0,6624	0,6833	0,6842	0,6222	0,6249	0,7882	0,6457	0,6818
Příznaky	age , trestbps , chol , restecg , exang , oldpeak , ca , slope_0 , thal_0								
Alfa = 1	0,8372	0,7442	0,8167	0,8684	0,7272	0,7000	0,8000	0,8947	0,6363
Příznaky	age , sex , fbs , exang , oldpeak , ca , cp_4 , cp_2 , slope_1								
Alfa = 1,5	0,8372	0,7907	0,8500	0,8947	0,7727	0,7143	0,8000	0,8684	0,6818
Příznaky	age , sex , trestbps , chol , fbs , exang , oldpeak , ca , cp_4 , cp_2 , slope_2 , slope_1 , slope_0 , thal_0 , thal_2								
Parametry	100 jedinců								
Alfa = 0,5	0,6222	0,6363	0,7333	0,7895	0,6363	0,6047	0,7167	0,7895	0,5910
Příznaky	sex , oldpeak								
Alfa = 1	0,6784	0,6666	0,7000	0,7465	0,7400	0,6822	0,7488	0,8000	0,4200
Příznaky	sex , ca , slope_2 , slope_1								
Alfa = 1,5	0,8372	0,7111	0,7833	0,8158	0,7272	0,7910	0,8500	0,8948	0,7727
Příznaky	age , sex , trestbps , chol , fbs , restecg , exang , oldpeak , ca , cp_4 , cp_2 , slope_2 , slope_1								

Tab. E.5: Výsledky HHO pro Dataset srdečního selhání

DATASET SRDEČNÍHO SELHÁNÍ									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,7000	0,6250	0,6889	0,9600	0,4000	0,6857	0,8070	0,8979	0,5889
Příznaky	ejection_fraction , platelets , trestbps , serum_sodium , time								
Alfa = 1	0,8333	0,7444	0,8070	0,9722	0,5238	0,7428	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , serum_sodium , smoking , time								
Alfa = 1,5	0,8000	0,7200	0,8246	0,9800	0,3480	0,7000	0,8224	0,9167	0,5714
Příznaky	ejection_fraction , high_blood_pressure , serum_sodium , time								
Parametry	30 jedinců								
Alfa = 0,5	0,7805	0,6061	0,7719	0,9444	0,4762	0,7805	0,8421	0,8888	0,7619
Příznaky	ejection_fraction , serum_creatinine , smoking , time								
Alfa = 1	0,7442	0,6850	0,7000	0,9600	0,4444	0,6857	0,8222	0,8667	0,5889
Příznaky	ejection_fraction , platelets , trestbps , serum_sodium , time								
Alfa = 1,5	0,8333	0,6666	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , platelets , serum_sodium , smoking , time								
Parametry	60 jedinců								
Alfa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age , ejection_fraction , high_blood_pressure , serum_sodium , smoking , time								
Alfa = 1	0,7805	0,6061	0,7719	0,9444	0,4762	0,7805	0,8421	0,8888	0,7619
Příznaky	ejection_fraction , serum_creatinine , smoking , time								
Alfa = 1,5	0,8544	0,6897	0,8000	0,8422	0,5238	0,7224	0,8421	0,8650	0,5800
Příznaky	age , ejection_fraction , platelets , serum_sodium , smoking , time								
Parametry	100 jedinců								
Alfa = 0,5	0,7500	0,6060	0,7720	0,9444	0,4762	0,5556	0,7200	0,8611	0,4762
Příznaky	anemia , diabetes , ejection_fraction , serum_creatinine , smoking , time								
Alfa = 1	0,7500	0,5161	0,7368	0,9444	0,3810	0,4848	0,7020	0,8889	0,3810
Příznaky	anemia , diabete , ejection_fraction , serum_creatinine , serum_sodium , sex , time								
Alfa = 1,5	0,8108	0,6286	0,7720	0,9167	0,5239	0,6667	0,7895	0,9167	0,5714
Příznaky	age , ejection_fraction , high_blood_pressure , serum_sodium , smoking , time								

Tab. E.6: Výsledky HHO pro Indický dataset

INDICKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,9832	0,9702	0,9650	0,9518	0,9744	0,9573	0,9500	0,9400	0,9573
Příznaky	serumcholesterol , maxheartrate, oldpeak, noofmajorvessels, chestpain_0 , chestpain_1 , chestpain_3 , slope_3 , slope_1 , slope_2 , slope_0 , restingrelectro_1, restingrelectro_0								
Alfa = 1	0,9957	0,9741	0,9700	0,9760	0,9658	0,9576	0,9500	0,9277	0,9658
Příznaky	restingBP , serumcholesterol , fastingbloodsugar , maxheartrate , exerciseangia , oldpeak , noofmajorvessels , chestpain_0 , chestpain_1 , chestpain_3 , slope_1 , slope_2 , slope_0 , restingrelectro_0 , restingrelectro_2								
Alfa = 1,5	0,9872	0,9569	0,9500	0,9518	0,9487	0,9661	0,9600	0,9400	0,9743
Příznaky	serumcholesterol, fastingbloodsugar , oldpeak , chestpain_0 , chestpain_1 , slope_2 , slope_0 , restingrelectro_2								
Parametry	30 jedinců								
Alfa = 0,5	0,9914	0,9784	0,9750	0,9880	0,9658	0,9702	0,9650	0,9518	0,9744
Příznaky	gender , restingBP , fastingbloodsugar , maxheartrate , chestpain_2 , chestpain_3 , slope_1 , slope_2 , slope_0 , restingrelectro_1 , restingrelectro_0 , restingrelectro_2								
Alfa = 1	0,9789	0,9572	0,9500	0,9398	0,9573	0,9451	0,9350	0,9036	0,9573
Příznaky	gender, serumcholesterol, fastingbloodsugar, oldpeak, noofmajorvessels, chestpain_1, slope_3 , slope_1, restingrelectro_1								
Alfa = 1,5	0,9832	0,9702	0,9650	0,9518	0,9744	0,9573	0,9500	0,9400	0,9573
Příznaky	serumcholesterol , maxheartrate, oldpeak, noofmajorvessels, chestpain_0 , chestpain_1 , chestpain_3 , slope_3 , slope_1 , slope_2 , slope_0 , restingrelectro_1 , restingrelectro_0								
Parametry	60 jedinců								
Alfa = 0,5	0,9915	0,9705	0,9650	0,9397	0,9829	0,9699	0,9650	0,9638	0,9658
Příznaky	gender , serumcholesterol , fastingbloodsugar , oldpeak , chestpain_2 , chestpain_3 , slope_3 , slope_1 , slope_0 , restingrelectro_2								
Alfa = 1	0,9832	0,9702	0,9650	0,9518	0,9746	0,9528	0,9450	0,9400	0,9450
Příznaky	gender , serumcholesterol , exerciseangia , oldpeak , noofmajorvessels , chestpain_0 , chestpain_1 , chestpain_3 , slope_1 , slope_2 , slope_0 , restingrelectro_0 , restingrelectro_2								
Alfa = 1,5	0,9872	0,9569	0,9500	0,9518	0,9487	0,9661	0,9600	0,9400	0,9743
Příznaky	serumcholesterol, fastingbloodsugar , oldpeak , chestpain_0 , chestpain_1 , slope_2 , slope_0 , restingrelectro_2								
Parametry	100 jedinců								
Alfa = 0,5	0,9957	0,9741	0,9700	0,9759	0,9658	0,9617	0,9550	0,9400	0,9658
Příznaky	gender, restingBP, fastingbloodsugar , maxheartrate, chestpain_0, chestpain_1 , slope_3, slope_1 , restingrelectro_2								
Alfa = 1	0,9915	0,9828	0,9800	0,9880	0,9744	0,9661	0,9600	0,9398	0,9743
Příznaky	age , gender , restingBP , fastingbloodsugar , maxheartrate , oldpeak , chestpain_2 , chestpain_0 , slope_3 , slope_1 , slope_0 , restingrelectro_2								
Alfa = 1,5	0,9915	0,9620	0,9475	0,9724	0,9444	0,9587	0,9520	0,9356	0,9700
Příznaky	age , gender , restingBP , fastingbloodsugar , maxheartrate , oldpeak , chestpain_2 , chestpain_0 , slope_3 , slope_1 , restingrelectro_2								

Tab. E.7: Výsledky FPA pro Klevelandský dataset

KLEVELANDSKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,8889	0,8085	0,8500	0,8421	0,8626	0,6957	0,7667	0,7895	0,7273
Příznaky	age, sex, trestbps, chol, fbs, restecg, oldpeak, ca, cp_4, slope_1, thal_1, thal_0, thal_2								
Alfa = 1	0,8261	0,8261	0,8667	0,8684	0,8636	0,7170	0,7500	0,6842	0,8636
Příznaky	restecg, ca, cp_4, slope_1, thal_1								
Alfa = 1,5	0,8372	0,8182	0,8667	0,8947	0,8182	0,7317	0,8167	0,8947	0,6818
Příznaky	age, sex, chol, fbs, restecg, thalach, ca, cp_2, slope_2, slope_1, slope_0, thal_1, thal_2								
Parametry	30 jedinců								
Alfa = 0,5	0,8837	0,8889	0,9167	0,9211	0,9091	0,6667	0,7500	0,7895	0,6818
Příznaky	age, sex, trestbps, chol, restecg, thalach, oldpeak, ca, cp_1, cp_4, slope_1, slope_0, thal_1, thal_2								
Alfa = 1	0,8571	0,8000	0,8500	0,8684	0,8182	0,7805	0,8500	0,9211	0,7273
Příznaky	age, chol, fbs, restecg, exang, oldpeak, ca, cp_2, slope_1, thal_1								
Alfa = 1,5	0,8636	0,8000	0,8500	0,8684	0,8182	0,7805	0,8500	0,9211	0,7273
Příznaky	chol, fbs, restecg, thalach, oldpeak, ca, cp_4, cp_2, slope_1, slope_0, thal_0, thal_2								
Parametry	60 jedinců								
Alfa = 0,5	0,8837	0,7727	0,8333	0,8684	0,7727	0,7143	0,8000	0,8684	0,6818
Příznaky	sex, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1, thal_0, thal_2								
Alfa = 1	0,8837	0,7727	0,8333	0,8684	0,7727	0,7143	0,8000	0,8684	0,6818
Příznaky	sex, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1, thal_0, thal_2								
Alfa = 1,5	0,8571	0,8700	0,9000	0,8947	0,9090	0,8261	0,8667	0,8684	0,8636
Příznaky	sex, trestbps, fbs, restecg, thalach, oldpeak, cp_4, slope_2, slope_1, thal_1								
Parametry	100 jedinců								
Alfa = 0,5	0,9268	0,8000	0,8500	0,8684	0,8182	0,7317	0,8167	0,8947	0,6818
Příznaky	sex, chol, fbs, restecg, thalach, exang, oldpeak, ca, cp_2, slope_1, slope_0, thal_1, thal_0								
Alfa = 1	0,8837	0,7727	0,8333	0,8684	0,7727	0,7692	0,8500	0,9474	0,6818
Příznaky	age, sex, fbs, restecg, thalach, exang, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1, thal_2								
Alfa = 1,5	0,8500	0,8444	0,8833	0,8947	0,8636	0,8182	0,8667	0,8947	0,8182
Příznaky	sex, restecg, thalach, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1								

Tab. E.8: Výsledky FPA pro Dataset srdečního selhání

DATASET SRDEČNÍHO SELHÁNÍ									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	0,8000	0,6667	0,8070	0,9722	0,5238	0,6667	0,7895	0,9167	0,5714
Příznaky	diabetes, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1	0,7805	0,5806	0,7719	0,9722	0,4286	0,7568	0,8421	0,9444	0,6667
Příznaky	age, anemia, ejection_fraction, time								
Alfa = 1,5	0,7568	0,6667	0,8070	0,9722	0,5238	0,5806	0,7719	0,9722	0,4286
Příznaky	age, diabetes, ejection_fraction, platelets, serum_sodium, smoking, time								
Parametry	30 jedinců								
Alfa = 0,5	0,8205	0,7059	0,8246	0,9722	0,5714	0,6286	0,7719	0,9167	0,5238
Příznaky	anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								
Alfa = 1	0,8108	0,6471	0,7895	0,9444	0,5238	0,4667	0,7193	0,9444	0,3333
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, smoking, time								
Alfa = 1,5	0,8205	0,6471	0,7895	0,9444	0,5239	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Parametry	60 jedinců								
Alfa = 0,5	0,8108	0,6471	0,7895	0,9444	0,5238	0,4667	0,7193	0,9444	0,3333
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, smoking, time								
Alfa = 1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5	0,8108	0,6286	0,7719	0,9167	0,5238	0,3571	0,6842	0,9444	0,2381
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, smoking, time								
Parametry	100 jedinců								
Alfa = 0,5	0,8205	0,6471	0,7895	0,9444	0,5239	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Alfa = 1	0,8108	0,6286	0,7719	0,9167	0,5238	0,3571	0,6842	0,9444	0,2381
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, smoking, time								
Alfa = 1,5	0,7692	0,5625	0,7549	0,9444	0,4286	0,4848	0,7018	0,8889	0,3810
Příznaky	anemia, creatinine_phosphokinase, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								

Tab. E.9: Výsledky FPA pro Indický dataset

INDICKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5	1,0000	0,9784	0,9750	0,9880	0,9658	0,9664	0,9600	0,9277	0,9829
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheart, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, slope_0, restingrelectro_0, restingrelectro_2								
Alfa = 1	1,0000	0,9784	0,9750	0,9880	0,9658	0,9701	0,9650	0,9518	0,9744
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheart, chestpain_2, chestpain_0, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1,5	0,9914	0,9784	0,9700	0,9759	0,9658	0,9746	0,9700	0,9518	0,9829
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_1, slope_2, slope_0, restingrelectro_2								
Parametry	30 jedinců								
Alfa = 0,5	1,0000	0,9784	0,9700	0,9759	0,9658	0,9620	0,9550	0,9277	0,9744
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_1, slope_3, slope_2, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1	0,9957	0,9784	0,9750	0,9880	0,9658	0,9565	0,9500	0,9639	0,9401
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, exerciseangia, noofmajorvessels, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_2, slope_0, restingrelectro_0, restingrelectro_2								
Alfa = 1,5	0,9915	0,9741	0,9700	0,9759	0,9658	0,9492	0,9400	0,9156	0,9573
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_2, chestpain_0, chestpain_1, slope_3, slope_1, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Parametry	60 jedinců								
Alfa = 0,5	0,9957	0,9784	0,9750	0,9880	0,9658	0,9532	0,9450	0,9277	0,9573
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, exerciseangia, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1	0,9957	0,9784	0,9700	0,9759	0,9658	0,9702	0,9650	0,9518	0,9744
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, restingrelectro_1, restingrelectro_2								
Alfa = 1,5	0,9915	0,9787	0,9750	0,9639	0,9829	0,9664	0,9600	0,9277	0,9829
Příznaky	age, gender, restingBP, serumcholesterol, maxheartrate, exerciseangia, oldpeak, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_0, restingrelectro_2								
Parametry	100 jedinců								
Alfa = 0,5	1,0000	0,9784	0,9750	0,9759	0,9744	0,9528	0,9450	0,9398	0,9487
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, slope_3, slope_2, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1	1,0000	0,9741	0,9700	0,9759	0,9658	0,9664	0,9600	0,9277	0,9829
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, chestpain_2, chestpain_0, chestpain_1, slope_3, slope_1, restingrelectro_1, restingrelectro_0								
Alfa = 1,5	1,0000	0,9787	0,9750	0,9639	0,9829	0,9664	0,9600	0,9277	0,9829
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_1, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								

Tab. E.10: Výsledky CSO pro Klevelandský dataset

KLEVELANDSKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5, pa = 0,1	0,9048	0,8000	0,8500	0,8684	0,8181	0,7143	0,8000	0,8684	0,6818
Príznaky	age, sex, fbs, restecg, thalach, exang, ca, cp_1, cp_2, slope_2, slope_1, slope_0, thal_2								
Alfa = 0,5, pa = 0,25	0,9048	0,8095	0,8667	0,9211	0,7727	0,7317	0,8167	0,8947	0,6818
Príznaky	age, sex, thalach, exang, olpeak, ca, cp_2, slope_2, slope_1, thal_1, thal_2								
Alfa = 0,5, pa = 0,5	0,9048	0,8293	0,8833	0,9474	0,7727	0,7273	0,8000	0,8421	0,7273
Príznaky	sex, chol, thalach, exang, olpeak, ca, cp_1, cp_2, slope_1, thal_2								
Alfa = 1, pa = 0,1	0,9048	0,8000	0,8500	0,8684	0,8182	0,7619	0,8333	0,8947	0,7273
Príznaky	age, fbc, restecg, exang, ca, cp_1, cp_2, slope_2, slope_1, slope_0, thal_1								
Alfa = 1, pa = 0,25	0,8837	0,8182	0,8667	0,8947	0,8182	0,7556	0,8167	0,8421	0,7727
Príznaky	sex, trestbps, chol, restecg, thalach, ca, cp_2, slope_2, slope_1, slope_0, thal_1								
Alfa = 1, pa = 0,5	0,9048	0,8000	0,8500	0,8684	0,8182	0,8000	0,8500	0,8684	0,8182
Príznaky	age, sex, chol, fbc, restecg, oldpeak, ca, cp_1, cp_2, slope_2, slope_1, slope_0, thal_1								
Alfa = 1,5, pa = 0,1	0,8696	0,8444	0,8833	0,8947	0,8636	0,7083	0,7667	0,7632	0,7727
Príznaky	age, sex, trestbps, chol, fbc, restecg, thalach, oldpeak, ca, cp_4, slope_2, thal_1, thal_0, thal_2								
Alfa = 1,5, pa = 0,25	0,9048	0,8571	0,9000	0,9474	0,8182	0,7556	0,8167	0,8421	0,7727
Príznaky	sex, trestbps, chol, fbc, restecg, thalach, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, thal_2								
Alfa = 1,5, pa = 0,5	0,9048	0,8000	0,8500	0,8684	0,8182	0,7500	0,8333	0,9211	0,8618
Príznaky	age, sex, fbc, restecg, thalach, exang, ca, cp_2, slope_1, slope_0, thal_1, thal_0								
Parametry	30 jedinců								
Alfa = 0,5, pa = 0,1	0,8837	0,8000	0,8500	0,8684	0,8182	0,7317	0,8167	0,8947	0,6818
Príznaky	age, sex, fbc, restecg, thalach, exang, ca, cp_1, cp_2, slope_2, slope_0, thal_1, thal_0								
Alfa = 0,5, pa = 0,25	0,7442	0,7442	0,8167	0,8684	0,7273	0,7111	0,7833	0,8158	0,7273
Príznaky	chol, restecg, thalach, exang, oldpeak, ca, cp_3, cp_2, slope_2, slope_0, thal_1, thal_0								
Alfa = 0,5, pa = 0,5	0,9091	0,8182	0,8667	0,8947	0,8182	0,7727	0,8333	0,8684	0,7727
Príznaky	sex, chol, restecg, thalach, ca, cp_2, slope_2, slope_1, slope_0, thal_1, thal_0								
Alfa = 1, pa = 0,1	0,8837	0,8182	0,8667	0,8947	0,8182	0,7317	0,8167	0,8947	0,6818
Príznaky	sex, chol, fbc, restecg, thalach, ca, cp_2, slope_2, slope_1, slope_0, thal_1, thal_2								
Alfa = 1, pa = 0,25	0,9000	0,8000	0,8500	0,8684	0,8182	0,6667	0,7667	0,8421	0,6364
Príznaky	age, fbc, restecg, exang, oldpeak, ca, cp_1, cp_2, slope_2, thal_0, thal_2								
Alfa = 1, pa = 0,5	0,9048	0,8000	0,8500	0,8684	0,8182	0,8000	0,8500	0,8684	0,8182
Príznaky	age, sex, chol, fbc, restecg, oldpeak, ca, cp_1, cp_2, slope_2, slope_1, slope_0, thal_1								
Alfa = 1,5, pa = 0,1	0,8780	0,8000	0,8500	0,8684	0,8182	0,8293	0,8833	0,9474	0,7727
Príznaky	sex, trestbps, restecg, thalach, oldpeak, cp_1, cp_4, slope_2, thal_1								
Alfa = 1,5, pa = 0,25	0,8837	0,8444	0,8833	0,8947	0,8636	0,7442	0,8167	0,8684	0,7273
Príznaky	sex, trestbps, fbc, restecg, thalach, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1, thal_2								
Alfa = 1,5, pa = 0,5	0,9000	0,7143	0,8000	0,8684	0,6818	0,7500	0,8333	0,9211	0,6818
Príznaky	sex, chol, restecg, thalach, exang, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1, thal_0								

Tab. E.11: Výsledky CSO pro Klevelandský dataset

KLEVELANDSKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	60 jedinců								
Alfa = 0,5, pa = 0,1	0,9268	0,8000	0,8500	0,8684	0,8182	0,7317	0,8167	0,8947	0,6818
Príznaky	sex, chol, fbs, restecg, thalach, exang, oldpeak, ca, cp_2, slope_1, slope_0, thal_1, thal_0								
Alfa = 0,5, pa = 0,25	0,9000	0,8000	0,8500	0,8684	0,8182	0,7500	0,8333	0,9211	0,6818
Príznaky	age, sex, restecg, thalach, exang, oldpeak, ca, cp_2, slope_2, slope_1, thal_1, thal_0, thal_2								
Alfa = 0,5, pa = 0,5	0,9048	0,8095	0,8667	0,9211	0,7727	0,7391	0,8000	0,8158	0,7727
Príznaky	sex, trestbps, restecg, thalach, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1, thal_2								
Alfa = 1, pa = 0,1	0,8889	0,8261	0,8667	0,8684	0,8636	0,7907	0,8500	0,8947	0,7727
Príznaky	sex, chol, restecg, thalach, oldpeak, cp_1, slope_2, thal_1								
Alfa = 1, pa = 0,25	0,8837	0,8000	0,8500	0,8684	0,8182	0,7143	0,8000	0,8684	0,8182
Príznaky	age, sex, trestbps, fbs, restecg, thalach, exang, ca, cp_2, slope_1, slope_0, thal_1, thal_0								
Alfa = 1, pa = 0,5	0,9048	0,7727	0,8333	0,8684	0,7727	0,7556	0,8167	0,8421	0,7727
Príznaky	sex, trestbps, restecg, thalach, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1, thal_2								
Alfa = 1,5, pa = 0,1	0,8837	0,8500	0,9000	0,9737	0,7727	0,7442	0,8167	0,8684	0,7273
Príznaky	age, sex, trestbps, chol, oldpeak, ca, cp_1, cp_2, slope_0, thal_1, thal_0								
Alfa = 1,5, pa = 0,25	0,9048	0,8181	0,8666	0,8947	0,8181	0,7907	0,8500	0,8947	0,7727
Príznaky	sex, trestbps, chol, fbs, restecg, thalach, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, thal_2								
Alfa = 1,5, pa = 0,5	0,8780	0,7826	0,8333	0,8421	0,8181	0,7442	0,8166	0,8684	0,7272
Príznaky	age, sex, trestbps, fbs, restecg, ca, cp_2, slope_2, slope_1, slope_0, thal_1								
Parametry	100 jedinců								
Alfa = 0,5, pa = 0,1	0,8636	0,8181	0,8667	0,8947	0,8181	0,7272	0,8000	0,8421	0,7272
Príznaky	sex, chol, restecg, thalach, oldpeak, ca, cp_2, slope_1, slope_0, thal_1, thal_0, thal_2								
Alfa = 0,5, pa = 0,25	0,8780	0,7907	0,8500	0,8947	0,7727	0,7727	0,8333	0,8684	0,7727
Príznaky	age, sex, restecg, oldpeak, ca, cp_1, cp_3, cp_2, slope_2, slope_1, thal_1								
Alfa = 0,5, pa = 0,5	0,8837	0,7441	0,8167	0,8684	0,7272	0,7907	0,8500	0,8947	0,7727
Príznaky	age, sex, fbs, restecg, exang, ca, cp_2, slope_2, slope_1								
Alfa = 1, pa = 0,1	0,9000	0,7441	0,8166	0,8684	0,7272	0,7179	0,8166	0,9220	0,6363
Príznaky	sex, restecg, thalach, exang, oldpeak, ca, cp_4, cp2, slope_2, slope_1, thal_1, thal_0								
Alfa = 1, pa = 0,25	0,8636	0,8095	0,8666	0,9211	0,7727	0,7442	0,8167	0,8684	0,7272
Príznaky	sex, trestbps, chol, restecg, thalach, exang, oldpeak, ca, cp_1, cp_4, cp2, slope_2, slope_1, slope_0, thal_2								
Alfa = 1, pa = 0,5	0,9000	0,8000	0,8500	0,8684	0,8181	0,7499	0,8333	0,9211	0,6818
Príznaky	sex, restecg, thalach, exang, oldpeak, ca, cp2, slope_1, thal_1, thal_2								
Alfa = 1,5, pa = 0,1	0,8837	0,8372	0,8833	0,9211	0,8181	0,8000	0,8500	0,8684	0,8181
Príznaky	sex, trestbps, chol, fbs, restecg, thalach, oldpeak, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1								
Alfa = 1,5, pa = 0,25	0,8889	0,8261	0,8667	0,8684	0,8636	0,7907	0,8500	0,8947	0,7727
Príznaky	sex, chol, restecg, thalach, oldpeak, cp_1, slope_2, thal_1								
Alfa = 1,5, pa = 0,5	0,8889	0,8085	0,8500	0,8421	0,8636	0,6956	0,7667	0,7895	0,7273
Príznaky	age, sex, trestbps, chol, fbs, restecg, oldpeak, ca, cp_4, slope_1, thal_1, thal_0, thal_2								

Tab. E.12: Výsledky CSO pro Dataset srdečního selhání

DATASET SRDEČNÍHO SELHÁNÍ									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry									
10 jedinců									
Alfa = 0,5, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,5	0,8205	0,6471	0,7895	0,9444	0,5238	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Alfa = 1, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,1	0,8205	0,7059	0,8246	0,9722	0,5714	0,6286	0,7719	0,9167	0,5238
Příznaky	anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								
Alfa = 1,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Parametry									
30 jedinců									
Alfa = 0,5, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,5	0,8205	0,6471	0,7895	0,9444	0,5238	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Alfa = 1, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,1	0,8205	0,7059	0,8246	0,9722	0,5714	0,6286	0,7719	0,9167	0,5238
Příznaky	anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								
Alfa = 1,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								

Tab. E.13: Výsledky CSO pro Dataset srdečního selhání

DATASET SRDEČNÍHO SELHÁNÍ									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	60 jedinců								
Alfa = 0,5, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,5	0,8205	0,6471	0,7895	0,9444	0,5238	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Alfa = 1, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,1	0,8205	0,7059	0,8246	0,9722	0,5714	0,6286	0,7719	0,9167	0,5238
Příznaky	anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								
Alfa = 1,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Parametry	100 jedinců								
Alfa = 0,5, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 0,5, pa = 0,5	0,8205	0,6471	0,7895	0,9444	0,5238	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Alfa = 1, pa = 0,1	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,1	0,8205	0,7059	0,8246	0,9722	0,5714	0,6286	0,7719	0,9167	0,5238
Příznaky	anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								
Alfa = 1,5, pa = 0,25	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								
Alfa = 1,5, pa = 0,5	0,8333	0,6667	0,8070	0,9722	0,5238	0,7429	0,8421	0,9722	0,6190
Příznaky	age, ejection_fraction, platelets, serum_sodium, smoking, time								

Tab. E.14: Výsledky CSO pro Indický dataset

INDICKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Alfa = 0,5, pa = 0,1	1,0000	0,9264	0,9150	0,9157	0,9145	0,9258	0,9150	0,9277	0,9060
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, chestpain_2, chestpain_0, chestpain_1, slope_0, restingrelectro_2								
Alfa = 0,5, pa = 0,25	1,0000	0,9740	0,9700	0,9759	0,9658	0,9658	0,9600	0,9518	0,9658
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_2, chestpain_0, slope_3, slope_1, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 0,5, pa = 0,5	1,0000	0,9829	0,9800	0,9759	0,9829	0,9705	0,9650	0,9398	0,9829
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_1, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1, pa = 0,1	1,0000	0,9829	0,9800	0,9880	0,9744	0,9705	0,9650	0,9398	0,9829
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, oldpeak, chestpain_2, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1, pa = 0,25	1,0000	0,9784	0,9750	0,9880	0,9658	0,9573	0,9500	0,9400	0,9573
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1, pa = 0,5	1,0000	0,9740	0,9700	0,9880	0,9573	0,9488	0,9400	0,9277	0,9487
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_2, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1,5, pa = 0,1	1,0000	0,9784	0,9750	0,9880	0,9658	0,9705	0,9650	0,9400	0,9830
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1,5, pa = 0,25	1,0000	0,9700	0,9650	0,9639	0,9658	0,9661	0,9600	0,9400	0,9744
Příznaky	age, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, restingrelectro_0, restingrelectro_2								
Alfa = 1,5, pa = 0,5	0,9871	0,9660	0,9600	0,9639	0,9573	0,9447	0,9350	0,9157	0,9487
Příznaky	age, restingBP, serumcholesterol, maxheartrate, chestpain_2, chestpain_1, chestpain_3, slope_3, slope_1, slope_2, restingrelectro_1, restingrelectro_2								
Parametry	30 jedinců								
Alfa = 0,5, pa = 0,1	1,0000	0,9784	0,9750	0,9880	0,9658	0,9748	0,9700	0,9518	0,9830
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_0, restingrelectro_2								
Alfa = 0,5, pa = 0,25	1,0000	0,9740	0,9700	0,9759	0,9658	0,9576	0,9500	0,9277	0,9658
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, restingrelectro_1, restingrelectro_2								
Alfa = 0,5, pa = 0,5	1,0000	0,9740	0,9700	0,9759	0,9658	0,9576	0,9500	0,9277	0,9658
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, restingrelectro_1, restingrelectro_2								
Alfa = 1, pa = 0,1	1,0000	0,9787	0,9750	0,9639	0,9829	0,9620	0,9550	0,9277	0,9744
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, slope_2, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1, pa = 0,25	1,0000	0,9700	0,9650	0,9639	0,9658	0,9569	0,9500	0,9518	0,9487
Příznaky	age, gender, restingBP, serumcholesterol, maxheartrate, exerciseangia, noofmajorvessels, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_2								
Alfa = 1, pa = 0,5	0,9697	0,9610	0,9550	0,9639	0,9487	0,9532	0,9450	0,9277	0,9573
Příznaky	gender, restingBP, fastingbloodsugar, maxheartrate, exerciseangia, chestpain_0, chestpain_1, slope_3, slope_2, slope_0, restingrelectro_1								
Alfa = 1,5, pa = 0,1	1,0000	0,9784	0,9750	0,9880	0,9658	0,9705	0,9650	0,9400	0,9830
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Alfa = 1,5, pa = 0,25	1,0000	0,9700	0,9650	0,9639	0,9658	0,9661	0,9600	0,9400	0,9744
Příznaky	age, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, restingrelectro_0, restingrelectro_2								
Alfa = 1,5, pa = 0,5	0,9871	0,9660	0,9600	0,9639	0,9573	0,9447	0,9350	0,9157	0,9487
Příznaky	age, restingBP, serumcholesterol, maxheartrate, chestpain_2, chestpain_1, chestpain_3, slope_3, slope_1, slope_2, restingrelectro_1, restingrelectro_2								

Tab. E.15: Výsledky CSO pro Indický dataset

INDICKÝ DATASET									
Parametry	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
60 jedinců									
Alfa = 0,5, pa = 0,1	1,0000	0,9658	0,9600	0,9518	0,9658	0,9831	0,9800	0,9638	0,9915
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_2, slope_0, restingelectro_2								
Alfa = 0,5, pa = 0,25	1,0000	0,9700	0,9650	0,9759	0,9573	0,9310	0,9200	0,9157	0,9231
Příznaky	age, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_2, slope_0, restingelectro_0, restingelectro_2								
Alfa = 0,5, pa = 0,5	1,0000	0,9741	0,9700	0,9759	0,9658	0,9748	0,9700	0,9398	0,9915
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, nofmajorvessels, chestpain_2, chestpain_0, slope_3, slope_1, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1, pa = 0,1	1,0000	0,9741	0,9700	0,9759	0,9658	0,9658	0,9600	0,9518	0,9658
Příznaky	age, restingBP, serumcholesterol, fastingbloodsugar, exerciseangia, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingelectro_2								
Alfa = 1, pa = 0,25	1,0000	0,9784	0,9750	0,9880	0,9658	0,9573	0,9500	0,9400	0,9573
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1, pa = 0,5	1,0000	0,9740	0,9700	0,9880	0,9573	0,9488	0,9400	0,9277	0,9487
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_2, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1,5, pa = 0,1	0,9871	0,9660	0,9600	0,9639	0,9573	0,9447	0,9350	0,9157	0,9487
Příznaky	age, restingBP, serumcholesterol, maxheartrate, chestpain_2, chestpain_1, chestpain_3, slope_3, slope_1, slope_2, restingelectro_1, restingelectro_2								
Alfa = 1,5, pa = 0,25	0,9447	0,9000	0,8800	0,8313	0,9145	0,9322	0,9200	0,8916	0,9402
Příznaky	gender, restingBP, maxheartrate, oldpeak, nofmajorvessels, chestpain_1, chestpain_3, slope_3, restingelectro_1								
Alfa = 1,5, pa = 0,5	0,9697	0,9610	0,9550	0,9639	0,9487	0,9532	0,9450	0,9277	0,9573
Příznaky	gender, restingBP, fastingbloodsugar, maxheartrate, exerciseangia, chestpain_0, chestpain_1, slope_3, slope_2, slope_0, restingelectro_1								
100 jedinců									
Alfa = 0,5, pa = 0,1	1,0000	0,9741	0,9700	0,9759	0,9658	0,9437	0,9350	0,9398	0,9316
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingelectro_0, restingelectro_2								
Alfa = 0,5, pa = 0,25	1,0000	0,9741	0,9700	0,9759	0,9658	0,9614	0,9550	0,9518	0,9573
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, restingelectro_2								
Alfa = 0,5, pa = 0,5	1,0000	0,9741	0,9700	0,9759	0,9658	0,9759	0,9700	0,9398	0,9915
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, nofmajorvessels, chestpain_2, chestpain_0, slope_3, slope_1, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1, pa = 0,1	1,0000	0,9787	0,9750	0,9639	0,9829	0,9620	0,9550	0,9277	0,9744
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, slope_2, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1, pa = 0,25	1,0000	0,9784	0,9750	0,9880	0,9658	0,9573	0,9500	0,9400	0,9573
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1, pa = 0,5	1,0000	0,9740	0,9700	0,9880	0,9573	0,9488	0,9400	0,9277	0,9487
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_2, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1,5, pa = 0,1	0,9871	0,9660	0,9600	0,9639	0,9573	0,9447	0,9350	0,9157	0,9487
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingelectro_1, restingelectro_0, restingelectro_2								
Alfa = 1,5, pa = 0,25	1,0000	0,9700	0,9650	0,9639	0,9658	0,9661	0,9600	0,9400	0,9744
Příznaky	age, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, restingelectro_0, restingelectro_2								
Alfa = 1,5, pa = 0,5	0,9871	0,9660	0,9600	0,9639	0,9573	0,9447	0,9350	0,9157	0,9487

Tab. E.16: Výsledky WOA pro Klevelandský dataset

KLEVELANDSKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specificita - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specificita - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Beta = 0	0,8571	0,8636	0,9000	0,9211	0,8936	0,7805	0,8500	0,9211	0,7273
Příznaky	sex, fbs, restecg, thalach, oldpeak, ca, cp_4, cp_2, slope_1, slope_0, thal_1, thal_0, thal_2								
Beta = 0,5	0,8500	0,8571	0,9000	0,9474	0,8182	0,7826	0,8333	0,8421	0,8182
Příznaky	age, sex, chol, fbs, thalach, exang, oldpeak, ca, cp_2, slope_2, slope_1, thal_1								
Beta = 1	0,9000	0,7143	0,8000	0,8684	0,6818	0,7500	0,8333	0,9211	0,6818
Příznaky	sex, chol, restecg, thalach, exang, oldpeak, ca, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1, thal_0								
Parametry	30 jedinců								
Beta = 0	0,8636	0,7907	0,8500	0,8947	0,7727	0,7273	0,8000	0,8421	0,7273
Příznaky	age, sex, trestbps, fbs, restecg, oldpeak, ca, cp_4, cp_2, slope_2, slope_0, thal_1, thal_0								
Beta = 0,5	0,8780	0,8444	0,8833	0,8947	0,8636	0,6977	0,7833	0,8421	0,6818
Příznaky	age, sex, trestbps, fbs, restecg, thalach, ca, cp_1, cp_2, slope_2, slope_1, thal_1, thal_2								
Beta = 1	0,8837	0,8444	0,8833	0,8947	0,8636	0,7805	0,8500	0,9210	0,7273
Příznaky	sex, chol, restecg, thalach, oldpeak, ca, cp_2, slope_1, thal_0, thal_2								
Parametry	60 jedinců								
Beta = 0	0,9048	0,8182	0,8667	0,8947	0,8182	0,6977	0,7833	0,8421	0,6818
Příznaky	sex, chol, restecg, thalach, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1, thal_0, thal_2								
Beta = 0,5	0,8571	0,7556	0,8167	0,8421	0,7727	0,7804	0,8500	0,9211	0,7273
Příznaky	age, sex, chol, fbs, restecg, exang, oldpeak, ca, cp_2, slope_2, slope_1, slope_0, thal_1, thal_0, thal_2								
Beta = 1	0,8373	0,7727	0,8333	0,8684	0,7727	0,7143	0,8000	0,8684	0,6818
Příznaky	age, sex, trestbps, chol, fbs, restecg, thalach, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1, thal_0, thal_2								
Parametry	100 jedinců								
Beta = 0	0,8780	0,7727	0,8333	0,8684	0,7727	0,7442	0,8167	0,8684	0,7273
Příznaky	age, sex, restecg, exang, oldpeak, ca, cp_1, cp_2, slope_2, slope_1, slope_0, thal_1, thal_2								
Beta = 0,5	0,8500	0,8095	0,8667	0,9211	0,7727	0,7442	0,8167	0,8684	0,7273
Příznaky	age, sex, trestbps, chol, fbs, thalach, exang, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, slope_0, thal_1, thal_0								
Beta = 1	0,8837	0,7727	0,8333	0,8684	0,7727	0,7143	0,8000	0,8684	0,6818
Příznaky	sex, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, ca, cp_1, cp_4, cp_2, slope_2, slope_1, thal_1, thal_0, thal_2								

Tab. E.17: Výsledky WOA pro Dataset srdečního selhání

DATASET SRDEČNÍHO SELHÁNÍ									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Beta = 0	0,8108	0,6250	0,7895	0,9722	0,4762	0,6060	0,7719	0,9444	0,4762
Příznaky	anemia, ejection_fraction, platelets, serum_sodium, sex, time								
Beta = 0,5	0,7778	0,6667	0,8070	0,9722	0,5238	0,4706	0,6842	0,8611	0,3810
Příznaky	age, diabetes, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Beta = 1	0,8108	0,6471	0,7895	0,9444	0,5238	0,4667	0,7193	0,9444	0,3333
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, smoking, time								
Parametry	30 jedinců								
Beta = 0	0,7805	0,6060	0,7719	0,9444	0,4762	0,7805	0,8421	0,8889	0,7619
Příznaky	ejection_fraction, serum_creatinine, time								
Beta = 0,5	0,7368	0,4828	0,7368	0,9722	0,3333	0,4333	0,6667	0,9444	0,1905
Příznaky	anemia, diabetes, high_blood_pressure, platelets, serum_creatinine, serum_sodium, sex, smoking, time								
Beta = 1	0,8108	0,6471	0,7895	0,9444	0,5238	0,4667	0,7193	0,9444	0,3333
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, smoking, time								
Parametry	60 jedinců								
Beta = 0	0,8205	0,7059	0,8246	0,9722	0,5714	0,6286	0,7719	0,9167	0,5238
Příznaky	anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, time								
Beta = 0,5	0,8108	0,6471	0,7895	0,9444	0,5238	0,4667	0,7193	0,9444	0,3333
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, smoking, time								
Beta = 1	0,7894	0,7059	0,8247	0,9722	0,5714	0,3704	0,7018	0,9722	0,2381
Příznaky	anemia, diabetes, ejection_fraction, high_blood_pressure, platelets, serum_sodium, smoking, time								
Parametry	100 jedinců								
Beta = 0	0,8108	0,6250	0,7895	0,9722	0,4762	0,6060	0,7719	0,9444	0,4762
Příznaky	anemia, ejection_fraction, platelets, serum_sodium, smoking, time								
Beta = 0,5	0,8205	0,6471	0,7895	0,9444	0,5238	0,3871	0,6667	0,8889	0,2857
Příznaky	age, anemia, ejection_fraction, high_blood_pressure, platelets, serum_sodium, sex, time								
Beta = 1	0,8108	0,6250	0,7895	0,9722	0,4762	0,5333	0,7544	0,9722	0,3810
Příznaky	age, anemia, ejection_fraction, platelets, serum_sodium, smoking, time								

Tab. E.18: Výsledky WOA pro Indický dataset

INDICKÝ DATASET									
	Fitness hodnota	F1 - LR	Přesnost - LR	Specifická - LR	Senzitivita - LR	F1 - KNN	Přesnost - KNN	Specifická - KNN	Senzitivita - KNN
Parametry	10 jedinců								
Beta = 0	0,9957	0,9658	0,9600	0,9518	0,9658	0,9576	0,9500	0,9277	0,9658
Příznaky	restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, noofmajorvessels, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_2, slope_0, restingrelectro_1, restingrelectro_2								
Beta = 0,5	1,0000	0,9784	0,9750	0,9880	0,9658	0,9746	0,9700	0,9518	0,9829
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0								
Beta = 1	1,0000	0,9784	0,9750	0,9880	0,9658	0,9702	0,9650	0,9518	0,9743
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_1, chestpain_3, slope_3, slope_1, restingrelectro_0, restingrelectro_2								
Parametry	30 jedinců								
Beta = 0	1,0000	0,9784	0,9750	0,9880	0,9658	0,9483	0,9400	0,9398	0,9401
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_2								
Beta = 0,5	0,9957	0,9697	0,9650	0,9759	0,9573	0,9528	0,9450	0,9398	0,9487
Příznaky	restingBP, serumcholesterol, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_1, chestpain_3, slope_1, slope_2, slope_0, restingrelectro_0, restingrelectro_2								
Beta = 1	0,9873	0,9741	0,9400	0,9759	0,9658	0,9620	0,9550	0,9277	0,9744
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_2, slope_0, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Parametry	60 jedinců								
Beta = 0	1,0000	0,9427	0,9350	0,9639	0,9145	0,9351	0,9250	0,9277	0,9231
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, chestpain_2, chestpain_0, chestpain_3, slope_1, slope_0, restingrelectro_2								
Beta = 0,5	0,9915	0,9700	0,9650	0,9639	0,9658	0,9710	0,9650	0,9157	1,0000
Příznaky	age, gender, serumcholesterol, fastingbloodsugar, maxheartrate, oldpeak, noofmajorvessels, slope_3, slope_1, slope_0, restingrelectro_0, restingrelectro_2								
Beta = 1	0,9829	0,9784	0,9750	0,9759	0,9744	0,9487	0,9400	0,9277	0,9487
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_2, chestpain_0, chestpain_1, chestpain_3, slope_3, slope_1, slope_0, restingrelectro_1, restingrelectro_0								
Parametry	100 jedinců								
Beta = 0	0,9915	0,9744	0,9700	0,9639	0,9744	0,9614	0,9550	0,9518	0,9573
Příznaky	gender, restingBP, serumcholesterol, maxheartrate, exerciseangia, oldpeak, noofmajorvessels, chestpain_2, chestpain_0, slope_3, slope_1, slope_2, slope_0, restingrelectro_0, restingrelectro_2								
Beta = 0,5	0,9957	0,9741	0,9700	0,9759	0,9658	0,9617	0,9550	0,9398	0,9658
Příznaky	age, gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, exerciseangia, noofmajorvessels, chestpain_2, chestpain_0, chestpain_3, slope_3, slope_1, slope_2, restingrelectro_1, restingrelectro_0, restingrelectro_2								
Beta = 1	1,0000	0,9741	0,9700	0,9759	0,9658	0,9702	0,9650	0,9277	0,9915
Příznaky	gender, restingBP, serumcholesterol, fastingbloodsugar, maxheartrate, noofmajorvessels, chestpain_2, chestpain_1, chestpain_3, slope_3, slope_1, restingrelectro_1, restingrelectro_0, restingrelectro_2								