



# 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

## DETEKCE STRESU A ÚNAVY V KOMPLEXNÍCH DATECH ŘIDIČE

STRESS AND FATIGUE DETECTION IN COMPLEX DRIVER'S DATA

### DIPLOMOVÁ PRÁCE

MASTER'S THESIS

### AUTOR PRÁCE

AUTHOR

Bc. Sabína Šimoňáková

### VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Martin Mézl, Ph.D.

BRNO 2021

# Diplomová práce

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

Ústav biomedicínského inženýrství

**Studentka:** Bc. Sabína Šimoňáková

**ID:** 195197

**Ročník:** 2

**Akademický rok:** 2020/21

**NÁZEV TÉMATU:**

## **Detekce stresu a únavy v komplexních datech řidiče**

**POKYNY PRO VYPRACOVÁNÍ:**

1) Proveďte literární rešerši metod pro detekci únavy během řízení auta s využitím dat z různých systémů. 2) Seznamte se s databází jízdních dat řidičů měřenou ve spolupráci s Ústavem soudního inženýrství VUT v Brně. Proveďte výběr vhodných úseků dat pro rozlišení únavy a stresu řidiče. 3) Navrhněte metodiku pro detekci únavy a stresu v těchto datech. Vybrané parametry z dat ověřte pomocí deskriptivní statistiky. 4) Vyberte vhodné kombinace příznaků, které podpoří metodiku. Implementujte výpočet těchto příznaků a následnou detekci. 5) Práci rozšiřte o srovnání různých detekčních algoritmů a pokuste se o automatickou detekci únavy a stresu. 6) Diskutujte výsledky práce s ohledem na publikované přístupy.

**DOPORUČENÁ LITERATURA:**

[1] NEMCOVA, Andrea, Martin SEITL, Tomas DOMINIK, et al. Multimodal Features for Detection of Driver Stress and Fatigue: Review. IEEE Transactions on Intelligent Transportation Systems. Early Access, March 2020. ISSN 1524-9050. DOI: 10.1109/TITS.2020.2977762.

[2] SIKANDER, Gulbadan a Shahzad ANWAR. Driver Fatigue Detection Systems: A Review. IEEE Transactions on Intelligent Transportation Systems. 2019, 20(6), 2339-2352. DOI: 10.1109/TITS.2018.2868499. ISSN 1524-9050.

**Termín zadání:** 8.2.2021

**Termín odevzdání:** 21.5.2021

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

**prof. Ing. Ivo 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**

Cieľom diplomovej práce je detekcia stresu a únavy z biologických signálov vodiča. V úvode sú opísané publikované metódy detekcie a oboznámenie sa s teoretickými znalosťami potrebnými k vypracovaniu danej témy. Prvým krokom v praktickej časti bola práca s databázou nameraných jázd a výber relevantných úsekov. Nasleduje extrakcia a selekcia príznakov. V práci bolo použitých 5 klasifikačných algoritmov na detekciu stresu a únavy a bola vykonaná predikcia na reálnych dátach. V poslednej časti je porovnanie najlepšieho modelu tejto práce s výsledkami v publikáciách.

## **Kľúčové slová**

Biologické signály, stres, únava, detekcia, databáza jázd, príznaky, klasifikačné modely, klasifikačná presnosť.

## **Abstract**

Main aim of our thesis is fatigue and stress detection from biological signals of a driver. Introduction contains information on published methods of detection and thoroughly informs readers about theoretical background necessary for our thesis. In the practical application we have firstly worked with a database of measured rides and subsequently chose their most relevant sections. Extraction and selection of features followed afterward. Five different classification models for tiredness and stress detection were used in the thesis and prediction was based on actual data. Lastly, the final section compares the best model of our thesis with the already published results.

## **Keywords**

Biological signals, stress, fatigue, detection, driving database, features, classification models, classification accuracy.

## **Bibliografická citácia**

ŠIMOŇÁKOVÁ, Sabína. *Detekce stresu a únavy v komplexních datech řidiče* [online]. Brno, 2021 [cit. 2021-05-17]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/134433>. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce Martin Mézl.

## Prehlásenie autora o pôvodnosti diela

**Meno a priezvisko študenta:** *Bc. Sabína Šimoňáková*

**VUT ID študenta:** *195197*

**Typ práce:** *Diplomová práca*

**Akademický rok:** *2020/21*

**Téma záverečnej práce:** *Detekce stresu a únavy v komplexních datech řidiče*

Prehlasujem, že svoju záverečnú prácu som vypracovala samostatne pod vedením vedúceho záverečnej práce a s použitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej záverečnej práce ďalej prehlasujem, že v súvislosti s vytvorením tejto záverečnej práce som neporušila autorské práva tretích osôb, najmä som nezasiahla nedovoleným spôsobom do cudzích autorských práv osobnostných a som si plne vedomá následkov porušenia ustanovenia § 11 a nasledujúcich autorského zákona č. 121/2000 Sb., vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákonníka č. 40/2009 Sb.

V Brne dňa: 3.5.2021

-----  
podpis autora

## **Pod'akovanie**

Ďakujem vedúcemu diplomovej práce Ing. Martin Mézl, Ph.D. za účinnú metodickú, pedagogickú a odbornú pomoc a ďalšie cenné rady pri spracovaní mojej diplomovej práce. Ďalej ďakujem svojej rodine a priateľovi za morálnu podporu počas celého štúdia a Daniele Ferencovej za pravopisnú a štylistickú kontrolu diplomovej práce.

V Brne dňa: 3.5.2021

-----  
podpis autora

# Obsah

<b>ZOZNAM OBRÁZKOV .....</b>	<b>10</b>
<b>ZOZNAM TABULIEK .....</b>	<b>11</b>
<b>ZOZNAM GRAFOV .....</b>	<b>13</b>
<b>ÚVOD .....</b>	<b>14</b>
<b>1. METÓDY PRE DETEKCIU ÚNAVY A STRESU .....</b>	<b>15</b>
1.1 DETEKCIA ZO SRDCOVÉHO SIGNÁLU .....	15
1.2 DETEKCIA Z EMG.....	17
1.3 DETEKCIA Z EDA.....	19
1.4 DETEKCIA Z RESPIRÁCIE.....	20
<b>2. SPRACOVANIE BIOLOGICKÝCH SIGNÁLOV .....</b>	<b>22</b>
2.1 SPRACOVANIE EKG .....	22
2.2 SPRACOVANIE EMG.....	22
2.3 SPRACOVANIE EDA .....	22
2.4 SPRACOVANIE SIGNÁLU RESPIRÁCIE .....	23
<b>3. VÝBER PRÍZNAKOV .....</b>	<b>24</b>
3.1 EXTRAKCIA PRÍZNAKOV .....	24
3.2 ZÁKLADNÁ ANALÝZA PRÍZNAKOV .....	24
3.3 MERANIE ZÁVISLOSTI MEDZI PRÍZNAKMI .....	24
3.4 SELEKCIA PRÍZNAKOV .....	26
<b>4. KLASIFIKAČNÉ METÓDY STROJOVÉHO UČENIA.....</b>	<b>28</b>
4.1 ROZHODOVACIE STROMY .....	28
4.2 NAIVNÝ BAYESOV KLASIFIKÁTOR.....	29
4.3 K – NAJBLIŽŠÍCH SUSEDOV .....	29
4.4 LOGISTICKÁ REGRESIA .....	29
4.5 SVM – METÓDA PODPORNÝCH VEKTOROV.....	30
<b>5. VYHODNOTENIE KLASIFIKÁTORA.....</b>	<b>31</b>
5.1 K-NÁSOBNÁ KRÍŽOVÁ VALIDÁCIA .....	31
5.2 KLASIFIKAČNÁ PRESNOŠŤ, MIERA CHYBY .....	32
5.3 MATICA NEURČITOSTI.....	32
5.4 ROC KRIVKA.....	33
5.5 F1 SKÓRE .....	34
5.6 VÝKON REGRESIE .....	34
<b>6. OPTIMALIZÁCIA HYPERPARAMETROV.....</b>	<b>35</b>
6.1 MANUÁLNE PREHLADÁVANIE .....	35
6.2 NÁHODNÉ PREHLADÁVANIE .....	35
6.3 VYHLADÁVACIA MRIEŽKA .....	36
6.4 BAYESOVSKÁ OPTIMALIZÁCIA .....	37
6.5 GENETICKÉ ALGORITMY.....	37

<b>7.</b>	<b>NAMERANÉ DÁTA Z JÁZD .....</b>	<b>38</b>
7.1	DATABÁZA.....	38
7.2	VIZUALIZÁCIA DÁT .....	40
<b>8.</b>	<b>VÝBER RELEVANTNÝCH DÁT.....</b>	<b>42</b>
8.1	DÁTA VYKAZUJÚCE STRES .....	42
8.2	DÁTA VYKAZUJÚCE ÚNAVU .....	42
<b>9.</b>	<b>SPRACOVANIE BIOSIGNÁLOV Z DATABÁZY .....</b>	<b>44</b>
9.1	NAČÍTANIE DÁT.....	44
9.2	PREDSPRACOVANIE DÁT.....	44
9.3	VÝSTUP PREDSPRACOVANIA .....	45
<b>10.</b>	<b>EXTRAKCIA RELEVANTNÝCH PRÍZNAKOV .....</b>	<b>46</b>
10.1	EKG.....	46
10.2	EMG .....	47
10.3	EDA.....	48
10.4	VÝSTUP EXTRAKCIE PRÍZNAKOV .....	48
<b>11.</b>	<b>ELIMINÁCIA DÁT A SELEKCIA PRÍZNAKOV.....</b>	<b>49</b>
11.1	ELIMINÁCIA NEVHODNÝCH DÁT .....	49
11.2	SELEKCIA PRÍZNAKOV PRE KLASIFIKÁCIU STRESU .....	49
11.3	VÝBER NAJLEPŠEJ SELEKČNEJ METÓDY PRE KLASIFIKÁCIU STRESU.....	51
11.4	SELEKCIA PRÍZNAKOV PRE KLASIFIKÁCIU ÚNAVY .....	52
11.5	VÝBER NAJLEPŠEJ SELEKČNEJ METÓDY PRE KLASIFIKÁCIU ÚNAVY .....	54
<b>12.</b>	<b>PRÍKLAD NAMERANÝCH BIOSIGNÁLOV .....</b>	<b>55</b>
12.1	ÚSEKY VYKAZUJÚCE NORMÁLNY STAV A STRESOVÝ STAV .....	55
12.2	ÚSEKY BEZ ÚNAVY A VYKAZUJÚCE ÚNAVU.....	59
<b>13.</b>	<b>OVERENIE POMOCOU POPISNEJ ŠTATISTIKY .....</b>	<b>64</b>
13.1	STRES .....	64
13.2	ÚNAVA .....	67
<b>14.</b>	<b>NÁVRH KLASIFIKAČNÝCH MODELOV A OPTIMALIZÁCIA HYPERPARAMETROV</b>	
	<b>70</b>	
14.1	ŠTANDARDIZÁCIA PRÍZNAKOV .....	70
14.2	ROZHODOVACIE STROMY .....	71
14.3	KLASIFIKÁTOR NAIVNÝ BAYES.....	71
14.4	METODA PODPORNÝCH VEKTOROV .....	72
14.5	K – NAJBLIŽŠÍCH SUSEDOV .....	73
14.6	LOGISTICKÁ REGRESIA .....	73
<b>15.</b>	<b>VÝSLEDKY KLASIFIKAČNÝCH MODELOV .....</b>	<b>74</b>
15.1	HODNOTENIE KLASIFIKAČNÝCH MODELOV PRE STRES.....	74
15.2	HODNOTENIE KLASIFIKAČNÝCH MODELOV PRE ÚNAVU.....	75
<b>16.</b>	<b>TESTOVANIE NA REÁLNYCH DÁTACH.....</b>	<b>77</b>
16.1	PREDIKCIA STRESU .....	77



16.2	PREDIKCIA ÚNAVY .....	79
<b>17.</b>	<b>POROVNANIE S PUBLIKÁCIAMI.....</b>	<b>80</b>
	<b>DISKUSIA .....</b>	<b>82</b>
	<b>ZÁVER .....</b>	<b>84</b>
	<b>LITERATÚRA.....</b>	<b>86</b>
	<b>ZOZNAM SYMBOLOV A SKRATIEK .....</b>	<b>93</b>
	<b>ZOZNAM PRÍLOH.....</b>	<b>95</b>

## ZOZNAM OBRÁZKOV

Obrázok 3. 1: Hughesov fenomén. Prevzaté z [41].	27
Obrázok 4. 1: Lineárna regresia vľavo, logistická regresia vpravo. Prevzaté z [31].	30
Obrázok 4. 2: Typy nadrovin označené zelenou priamkou (vľavo), optimálna nadrovina (vpravo), prevzaté z [33].	30
Obrázok 5. 1: Príklad K-násobnej krížovej validácie a výpočet presnosti, prevzaté z [39].	31
Obrázok 5. 2: Matica neurčitosti so zvýraznenými správne klasifikovanými hodnotami.	32
Obrázok 5. 3 - ROC krivka, sivá oblasť značí AUC, prevzaté z [38].	33
Obrázok 6. 1 : Náhodné prehľadávaný priestor vľavo, nájdené optimálnych hyperparametrov vpravo, prevzaté z [45].	36
Obrázok 6. 2 : Vyhľadávanie v mriežke vľavo, nájdené optimálnych hyperparametrov vpravo, prevzaté z [44].	37
Obrázok 7. 1 : Nasnímané polohy očí v jednom okamihu pomocou eyetrackera implementovaného v okuliaroch vodiča, naľavo ľavé oko, napravo pravé oko.	39
Obrázok 7. 2 : Pohľad na interiér vozidla z implementovanej kamery v okuliaroch vodiča. Značky s "QR kódmi". Červený terč.	40
Obrázok 7. 3 : Záznam signálov z rýchlostného pedálu a brzďového pedálu, trvanie 10 sekúnd.	40
Obrázok 7. 4 : Príklad spusteného programu pre konkrétne dáta so zobrazeným statickým obrazom z kamery, aktuálnou polohou na mape, signálmi z pedálov a biosignálov so synchronizáciou.	41
Obrázok 12. 1 : Biosignál EKG, trvanie 10s.	55
Obrázok 12. 2 : Biosignál EMG, trvanie 10s.	56
Obrázok 12. 3 : Biosignál EDA, trvanie 10s.	56
Obrázok 12. 4 : Biosignál EKG, 10s.	57
Obrázok 12. 5 : Biosignál EMG, 10s.	57
Obrázok 12. 6 : Biosignál EDA, 10s.	57
Obrázok 12. 7 : EKG biosignál, 10s.	58
Obrázok 12. 8 : EMG biosignál, 10s.	58
Obrázok 12. 9 : EKG biosignál, 10s.	59
Obrázok 12. 10 : EMG biosignál, 10s.	59
Obrázok 12. 11 : EDA biosignál - nevyhovujúci, 10s.	59
Obrázok 12. 12 : EKG biosignál, začiatok dennej jazdy, 10s.	60
Obrázok 12. 13 : EKG biosignál, koniec dennej jazdy, 10s.	60
Obrázok 12. 14 : EKG biosignál, koniec dennej jazdy, 10s.	61
Obrázok 12. 15 : EMG biosignál, koniec dennej jazdy, 10s.	61
Obrázok 12. 16 : EDA biosignál, začiatok dennej jazdy, 10s.	61
Obrázok 12. 17 : EDA biosignál, koniec dennej jazdy, 10s.	62
Obrázok 12. 18 : EKG biosignál, začiatok nočnej jazdy, 10s.	62
Obrázok 12. 19 : EKG biosignál, koniec nočnej jazdy, 10s.	63
Obrázok 12. 20 : EMG biosignál, začiatok nočnej jazdy, 10s.	63
Obrázok 12. 21 : EMG biosignál, koniec nočnej jazdy, 10s.	63
Obrázok 16. 1 : Úsek jazdy vykazujúci stres u vodiča – chodec vstúpil na vozovku z rady áut.	78
Obrázok 16. 2 : Záznam biosignálu EKG, u ktorého bol predikovaný stres.	78
Obrázok 16. 3 : Záznam biosignálu EMG, u ktorého bol predikovaný stres.	79

## ZOZNAM TABULIEK

Tabuľka 10. 1 : Extrahované príznaky zo signálu EKG. ....	46
Tabuľka 10. 2 : Príklady extrahovaných príznakov pre EKG. ....	47
Tabuľka 10. 3 : Extrahované príznaky zo signálu EMG. ....	47
Tabuľka 10. 4 : Príklady extrahovaných príznakov z EMG. ....	47
Tabuľka 10. 5 : Extrahované príznaky zo signálu EDA. ....	48
Tabuľka 10. 6 : Príklady extrahovaných príznakov z EDA. ....	48
Tabuľka 11. 1 : Prvých 8 príznakov s najvyššími váhami určenými algoritmom NCA pre stres. ....	51
Tabuľka 11. 2 : Prvých 8 príznakov s najvyššími váhami určenými algoritmom NCA pre únavu. ....	54
Tabuľka 13. 1 : Hodnoty parametrov analýzy EKG od všetkých subjektov z denného záznamu. Vysvetlivky : SH- stredná hodnota, SO – smerodajná odchýlka, R – rozptyl. ....	65
Tabuľka 13. 2 : Hodnoty parametrov analýzy EKG od všetkých subjektov z nočného záznamu. Vysvetlivky : SH- stredná hodnota, SO – smerodajná odchýlka, R – rozptyl. ....	65
Tabuľka 13. 3 : Hodnoty parametrov analýzy EMG od všetkých subjektov z denného záznamu. Vysvetlivky: mV2 – jednotka energie signálu. ....	65
Tabuľka 13. 4 : Hodnoty parametrov analýzy EMG od všetkých subjektov z nočného záznamu. Vysvetlivky: mV2 – jednotka energie signálu. ....	66
Tabuľka 13. 5 : Hodnoty parametrov analýzy EDA od všetkých subjektov z denného záznamu. ....	66
Tabuľka 13. 6 : Hodnoty parametrov analýzy EDA od všetkých subjektov z nočného záznamu. ....	66
Tabuľka 13. 7 : Hodnoty parametrov analýzy EKG zo začiatku jazdy a pred ukončením dennej jazdy pre všetky subjekty. ....	67
Tabuľka 13. 8 : Hodnoty parametrov analýzy EKG zo začiatku diaľnice a pred zjazdom z diaľnice počas nočnej jazdy pre všetky subjekty. ....	67
Tabuľka 13. 9 : Hodnoty parametrov analýzy EMG zo začiatku jazdy a pred ukončením dennej jazdy pre všetky subjekty. Vysvetlivky: mV2 – jednotka energie signálu. ....	68
Tabuľka 13. 10 : Hodnoty parametrov analýzy EMG zo začiatku diaľnice a pred zjazdom z diaľnice počas nočnej jazdy pre všetky subjekty. Vysvetlivky: mV2 – jednotka energie signálu. ....	68
Tabuľka 13. 11 : Hodnoty parametrov analýzy EDA zo začiatku jazdy a pred ukončením dennej jazdy pre všetky subjekty. ....	69
Tabuľka 13. 12 : Hodnoty parametrov analýzy EDA zo začiatku diaľnice a pred zjazdom z diaľnice počas nočnej jazdy pre všetky subjekty. ....	69
Tabuľka 14. 1 : Vybrané príznaky pre vstup do klasifikačných algoritmov. ....	70
Tabuľka 14. 2 : Tabuľka s vybranými optimalizovanými parametrami pre rozhodovací strom pre stres a únavu. ....	71
Tabuľka 14. 3 : Tabuľka s vybranými optimalizovanými parametrami pre klasifikátor Naivný Bayes pre stres a únavu. ....	72
Tabuľka 14. 4 : Tabuľka s vybranými optimalizovanými parametrami pre klasifikačnú metódu podporných vektorov pre stres a únavu. ....	72
Tabuľka 14. 5: Tabuľka s vybranými optimalizovanými parametrami pre klasifikačnú metódu K-najbližší sused pre stres a únavu. ....	73
Tabuľka 14. 6 : Tabuľka s vybranými parametrami pre klasifikačnú metódu logistická regresia pre stres a únavu. ....	73
Tabuľka 15. 1 : Výsledné zhodnotenie úspešnosti klasifikačných modelov pre stres. ....	74
Tabuľka 15. 2 : Výsledné zhodnotenie úspešnosti klasifikačných modelov pre únavu. ....	75

Tabuľka 16. 1 : Predikovaná trieda v 50 sekundovom zázname pre každý klasifikátor. ....	78
Tabuľka 17. 1 : Publikované metódy detekcie stresu. ....	80
Tabuľka 17. 2 : Publikované metódy detekcie únavy. ....	81

## ZOZNAM GRAFOV

<i>Graf 11. 1 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa algoritmu MRMR pre stres.</i>	50
<i>Graf 11. 2 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa chi-kvadrát testu pre stres.</i>	50
<i>Graf 11. 3 - Normalizované váhy príznakov udelených algoritmom NCA pre stres.</i>	51
<i>Graf 11. 4 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa algoritmu MRMR pre únavu.</i>	52
<i>Graf 11. 5 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa chi-kvadrát testu pre únavu.</i>	53
<i>Graf 11. 6 : Normalizované váhy príznakov udelených algoritmom NCA pre únavu.</i>	53
<i>Graf 15. 1 : ROC krivka ako výsledok hodnotenia klasifikačného modelu Naivný Bayes pre stres.</i>	75
<i>Graf 15. 2 : ROC krivka ako výsledok hodnotenia klasifikačného modelu Rozhodovací strom pre stres.</i>	76
<i>Graf 17. 1 : Porovnaná klasifikačná presnosť publikovaných detekčných algoritmov pre stres s najlepším klasifikačným modelom tejto práce.</i>	80
<i>Graf 17. 2 : Porovnaná klasifikačná presnosť publikovaných detekčných algoritmov pre únavu s najlepším klasifikačným modelom tejto práce.</i>	81

# ÚVOD

S rastúcou populáciou narastá počet automobilov a s tým súvisí zvýšený výskyt autonehôd, ktoré sú často spôsobené mikrosnávkou vodiča alebo pretrvávajúcou únavou. Ďalším aspektom, ktorý spôsobuje autonehody a nepozornosť vodiča sú stresové faktory, ktoré negatívne vplyvajú na psychiku vodiča. Moderné automobily majú v sebe implementované zariadenia, ktoré dokážu zaznamenávať aktivitu vodiča a kontrolovať či nedochádza k poklesu pozornosti a mikrosnávkou. Aj keď rôzne detekčné prístupy už boli komercializované a používajú ich výrobcovia automobilov, stále je potrebné vylepšovať výskum v tejto oblasti, aby bola zabezpečená čo najvyššia spoľahlivosť týchto výstražných systémov.

Monitorovanie stavu pozornosti sa považuje za jedno z najdôležitejších parametrov bezpečnej jazdy. Únava spomaľuje čas odozvy vodiča, čo zapríčiňuje elimináciu schopnosti vodiča jazdiť efektívne. Výskum v oblasti monitorovania vodiča nabral na obrátkach, špeciálne odhad pracovnej záťaže vodiča, identifikácie činnosti vodiča a rozpoznávanie štýlu jazdy.

Stres vyskytujúci sa počas vedenia vozidla je spojený so zhoršenými rozhodovacími schopnosťami, so zníženým výkonom a zhoršeným situačným povedomím. Preto je potrebné zlepšiť včasné odhalenie tejto situácie pre zlepšenie pozornosti vodiča. Boli vyvinuté rôzne systémy monitorujúce stres, s cieľom upozorniť vodiča na stav rizika. Pribúdajúca elektronika na palubnej doske a zvýšená koncentrácia áut na cestách si vyžaduje lepšie monitorovacie zariadenia. Analýza biologických signálov sa ukázala ako cenovo prijateľný a účinný spôsob detekcie rôznych fyziologických stavov vodiča.

Podľa prieskumu až 20 % nehôd v komerčnej doprave je zapríčinených únavou, 20 % smrteľných nehôd v Amerike bolo spôsobených ospalosťou vodiča a dokonca Pakistan zaregistroval 34 % dopravných nehôd ako dôsledok únavy [5]. Všetky štatistické údaje sú alarmujúce a vyžadujú si vysokú pozornosť výskumníkov v procese riešenia dopravných nehôd.

Táto diplomová práca sa venuje detekcii stresu a únavy ako dvom separátnym stavom, ktoré zvyšujú riziko dopravných nehôd. Jej výstupom je úspešnosť najlepšieho klasifikačného modelu strojového učenia, ktorý je otestovaný na reálnych dátach a porovnaný s dostupnými publikáciami.

# 1. METÓDY PRE DETEKCIU ÚNAVY A STRESU

Monitorovanie pozornosti vodiča je jeden z najdôležitejších parametrov pre bezpečnú jazdu. Záujem vodiča by mal byť počas celej jazdy sústredený na cestu a všetky faktory, ktoré negatívne ovplyvňujú riadenie, by mali byť v čo najväčšej miere eliminované. Častými faktormi, kvôli ktorým vodič stráca pozornosť, sú únava a stres.

Únava všeobecne spomaľuje odozvu človeka na akúkoľvek prijatú informáciu, čo môže spôsobovať komplikácie v doprave. Ospalosť a spánok sú faktory, ktoré každoročne prispievajú k tisíckam dopravných nehôd, zraneniam a smrteľným úrazom. Únavu vodiča je možné rozdeliť na dve podkategórie. Prvá je závislá na dennom čase, spánkovej deprivácii a predĺženým trvaním bdlosti a druhá závisí na vlastnostiach riadenia, ako sú splnenie pracovnej úlohy a jej trvanie [1]. Ďalšia klasifikácia môže byť na aktívnu, pasívnu a únavu súvisiacu so spánkom. Duševné vyčerpanie spôsobené niekoľkohodinovou prácou zapríčiňuje aktívnu únavu. Pasívna únava je spôsobená monotónnou jazdou. Vodič nemusí byť unavený a napriek tomu stráca pozornosť a zvyšuje sa jeho reakčný čas. Jedna štúdia uvádza, že jednotvárne cesty, ktorými sú diaľnice, zvyšujú pasívnu únavu u vodiča, čo má negatívny vplyv na počet nehôd odohraných na cestných komunikáciách [2].

Stres je normálna biologická reakcia na potencionálnu nebezpečnú situáciu. Respektíve je to odpoveď mozgu a tela na nejakú požiadavku. Stres sa prejavuje po psychickej stránke, v podobe úzkosti a depresie, alebo fyzickej stránke, ktorá ovplyvňuje činnosť väčšiny orgánov v tele [3]. U vodiča nie je vždy stres zapríčinený dopravnou situáciou, ale môže byť aj reakciou na rôznu životnú situáciu. Podobne môže byť stres spôsobený použitím rôznych omamných látok, zdravotným stavom a aj samotnou únavou, čo je spojené s potením rúk, zmenou srdcovej činnosti, krvného tlaku a dýchania.

Najlepšou voľbou získania vhodných dát pre detekciu únavy a stresu je nameranie biologických signálov pomocou dostupných senzorov. Medzi najdôležitejšie merania patrí EKG, PPG, HRV, EMG, EDA a respirácia. Sensory by mali splňovať rôzne kritéria, ako napríklad neinvazívnosť a pohodlnosť [4].

## 1.1 Detekcia zo srdcového signálu

Používajú sa rôzne spôsoby snímania signálov EKG a PPG. Meranie pomocou senzorov umiestnených na pokožke vodiča sa začali považovať za nie úplne vhodné, hlavne kvôli nekonformnosti. S postupujúcim vývojom došlo k implementácii senzorov do volantu a do bezpečnostného pásu vodiča, čo spôsobilo vyšší záujem u automobilových výrobcov. Pomerne vysoká chyba sa objavila u nameraných dátach, kde boli senzory umiestnené v sedadle vodiča, alebo v navlečenom páse [5].

Jednou z možností je použitie zabudovaných snímačov vo volante vozidla, ktoré boli navrhnuté na meranie nepretržitého signálu EKG z dlaní vodiča. V tomto prípade je zo signálu EKG analyzované HRV, ktoré patrí medzi neinvazívne techniky pre zhodnotenie autonómneho nervového systému(ANS), ktorý je ukazovateľom schopnosti človeka reagovať na každodenný tlak a stres z okolia. Neustálym meraním HRV je možné sledovať a indikovať nástup neuropatie ANS a takto minimalizovať možné rizikové faktory u vodiča, ktorými sú ospalosť, stres a srdcové choroby. Najčastejšie používané metódy analýzy HRV sú založené na časovej oblasti a frekvenčnej oblasti. Podstata HRV analýzy v časovej oblasti spočíva v intervaloch normal-to-normal(NN) medzi jednotlivými komplexami kontinuálneho EKG signálu a výpočtom vzoriek signálu. Na základe tohto merania sa vypočítajú štatistické parametre ako je priemerný tep srdca, štandardná odchýlka intervalu NN (SDNN), odmocnina priemerov kvadrátov diferencií susedných intervalov NN (RMSSD), ktoré indikujú rozdiely v nameraných NN intervaloch a umožňujú porovnanie HRV signálov počas únavy a ospalosti. Únava zapríčiňuje mierne zníženie priemerného tepu srdca, veľký pokles hodnôt SDNN a RMSSD. So začínajúcou ospalosťou sa prehĺbuje zníženie tepu srdca a mierne klesá SDNN aj RMSSD. Analýza vo frekvenčnej oblasti spočíva v meraní hustoty výkonného spektra použitím rýchlej Fourierovej transformácie(FFT), ktorá poskytuje informácie o troch spektrálnych zložkách, veľmi nízka frekvencia(VLF), nízka frekvencia(LF), vysoká frekvencia(HF), ktoré sú merané v absolútnych hodnotách. Rozsah frekvencie VLF je do 0,04 Hz, rozsah LF je 0,04 Hz – 0,15 Hz a HF je 0,18 Hz – 0,4 Hz. Distribúcia energie a stredná frekvencia LF a HF je premenlivá podľa zmien v ANS. Pomer zložiek LF a HF zdôrazňuje vyvážené správanie dvoch vetiev ANS. V tomto prípade hodnoty LF a HF s nadobúdajúcou únavou rapídne klesajú a mierny pokles nastáva v oboch zložkách pri ospalosti. Pre rozhodovanie o zdravotnom stave vodiča z hľadiska únavy, ospalosti alebo normálnej kondície, je najlepšou možnosťou pomer LF/HF. S narastajúcou únavou a ospalosťou tento pomer tiež narastá [6], podobne je tomu aj u stresu, kedy s narastajúcim stresom stúpa hodnota LF/HF [8].

Signál HRV je možné extrahovať aj zo signálu PPG, ktorého senzor je umiestnený vo volante vodiča. Metóda spočíva v prevedení extrakcie príznakov pomocou diskkrétnej vlnkovej transformácie (DWT), ktorá je založená na materskej vlnke Symlet 3. DWT dá rozklad signálu na množinu približných a podrobných koeficientov. Každý signál HRV je rozložený na osem úrovní s daným frekvenčným rozsahom. Štandardná najkratšia doba trvania pre analýzu pomeru LF/HF pre HRV je 2 minúty. Občas sa môžu použiť aj 1-minútové a 3-minútové. Výber príznakov prebieha pomocou ROC analýzy, ktorá pomáha zvoliť správne kritérium pre návrh klasifikátora. Vyberie sa pomer LF/HF a najlepšia vlnková funkcia s vysokou ROC oblasťou na vytvorenie príznakových vektorov pre tréningovanie podporných vektorov(SVM). Vyžadujú sa najmenej dve vektory príznakov a to stavová premenná a testovacia premenná, pričom stavová premenná nadobúda hodnoty 1-ospalá a -1- odpočínutá a hodnota z ROC analýzy od 0 po 1. Hodnota 0,5



ROC značí, že príznaky sú neoddeliteľné a hodnoty 0,7 a 0,3 vyjadrujú, že príznaky sú prijateľné pre klasifikáciu. SVM sa používa na automatickú klasifikáciu ospalosti pri vedení vozidla a to nájdením maximálnej vzdialenosti a maximálneho rozpätia medzi dvomi triedami. Výstupom SVM sú hodnoty -1, ktorá predstavuje odpočínutý stav vodiča a 1, ktorá predstavuje ospalý stav vodiča [7].

Jeden z ďalších možných prístupov je založený na evolučnom algoritme PSO, ktorý slúži na optimalizáciu hyperparametrov okna, ktorými sú veľkosť okna a stupeň prekrytia. Metóda sa delí na dve fázy. Prvá fáza pozostáva z predspracovania a druhá fáza zahŕňa samotnú optimalizáciu hyperparametrov okna. V prvom rade je signál EKG filtrovaný od šumu, následne sú extrahované R vrcholy a nakoniec je zmerané HRV. V druhej časti sa aplikuje samotný algoritmus PSO na nájdenie najvhodnejších hodnôt parametrov okna. Prebehne vyhodnotenie výkonu vyselektovaných hyperparametrických hodnôt a zostavenie modelu na detekciu stresu u vodiča. Vybrané hyperparametrické hodnoty sa použijú v modeli na detekciu stresu a na základe vybraných hodnôt sa vyhodnotí výkonnosť zostaveného modelu. Algoritmus PSO skúma rôzne hyperparametrické hodnoty v iteračnom procese, kým nenájde tie najvhodnejšie hodnoty. V každej iterácii je zostavený model detekcie úrovni stresu založený na EKG, ktorý obsahuje extrakciu a klasifikáciu príznakov. Proces vytvorenia modelu spočíva v segmentácii HRV dát na základe výberu veľkosti okna a stupňa prekrytia. Vypočítajú sa údaje ako je NN, SDNN, RMSSD, LF, HF, čiže údaje z časovej a frekvenčnej oblasti HRV. Celkový výkon sa považuje za ďalší príznak, ktorý sa používa na detekciu úrovne stresu u vodiča. Na klasifikáciu sa používa algoritmus Random Forest (RF), ktorý je kombináciou množstva náhodných stromov a každý klasifikátor predpovedá označenie triedy pre každý vstupný vektor. Stupeň prekrytia sa privádza do klasifikátora na tréning modelu a klasifikáciu úrovni stresu. Potom, pre každý vstupný vektor, sa označí trieda, ktorá je vybraná viackrát ako ostatné a bude označená za konečnú triedu. V poslednom kroku sa vyhodnotí výkonnosť vytvoreného modelu z hľadiska presnosti detekcie a účinnosti hyperparametrov pomocou 10-násobnej techniky krížovej validácie [9].

## 1.2 Detekcia z EMG

Medzi ďalšie biologické signály patrí signál EMG, ktorý môže byť taktiež použitý na detekciu únavy, poprípade aj stresu. Príznaky, ako je zmena priemernej frekvencie a strednej frekvencie, elektrická aktivita, stredná kvadratická hodnota amplitúdy a integrovaný elektromyogram sú extrahované z časovej alebo frekvenčnej oblasti povrchového EMG(sEMG) signálu a môžu predpovedať únavu [5].

FFT je najpopulárnejšia technika spracovania signálu pre analýzu sEMG. Inou metódou je vlnkový prístup na detekciu únavy, detekciu reakčného času alebo rozpoznanie vzoru pre signál sEMG. Niektoré štúdie uvádzajú, že použitie vlnkovej transformácie je lepšou alternatívou na analýzu sEMG ako FFT, pretože nedochádza

k stratám informácií v nestacionárnych podmienkach. V počiatočnej fáze je signál sEMG filtrovaný filtrom pásmová zadrž šiesteho rádu v hodnotách 15-500 Hz. Následne je signál filtrovaný filtrom Notch, 2. rádu v hodnotách 49-51 Hz. Jedná sa o tzv. filter odmietnutia pásma, ktorý oslabuje signály v konkrétnom frekvenčnom pásme a prepúšťa signály nad a pod týmto pásmom [13]. Na filtrovaný signál je aplikovaná diskretná vlnková transformácia(DWT) s vlnkou Daubechies(db5). Hodnoverný indikátor indexu únavy je zmena výkonu v frekvenčnom pásme 15-30 Hz. Pri vzorkovacej frekvencii 1000 Hz, piata úroveň rozkladu vlnky odpovedá frekvenčnému pásmu 15.6-31.3 Hz. Vývoj svalovej únavy je kontrolovaný amplitúdou aproximačných koeficientov. Sklon každého príznaku extrahovaného zo spektra signálu sEMG je ukazovateľ zmeny svalovej aktivity. Predovšetkým sa sleduje sklon priemerného výkonu z vlnkových koeficientov. V priebehu opakovania svalovej činnosti má kontraktilná sila tendenciu klesať. Takáto zmena svalovej aktivity je pozorovateľná ako posun dominantnej frekvencie signálu EMG k dolnému frekvenčnému pásmu spektra, čo sa považuje za indikátor nástupu únavy [14].

Ďalšou metódou na detekciu únavy zo signálu EMG je metóda súčtu sínusov. Najprv je signál filtrovaný filtrom pásmová zadrž, Butterworth v hodnotách 20-400 Hz. Bolo zistené, že krátke úseky sú vhodnejšie na analýzu nelineárnych časových úsekov signálu, preto sa extrahujú úseky pred a po únave v dĺžke trvania 250 ms. Tieto úseky signálu sa následne fitujú do nelineárnej krivky. Fitovanie krivky je technika pri ktorej sa vytvára krivka, ktorá najlepšie zodpovedá súboru hodnôt. Úmyslom je nájsť také koeficienty, ktoré najviac fitujú dáta. Zvyčajne sa používa nelineárna metóda Najmenších štvorcov, vďaka ktorej sa zníži suma chýb. Získajú sa dve dátové súbory, jeden menší súbor časového úseku a druhý produkovaný cez nafitovanú krivku, získaný pre unavený aj neunavený stav. Signál sEMG môže byť prezentovaný modelom sínusového signálu s amplitúdou, frekvenciou a fázou. Táto metodika zahŕňa metódu súčtu sínusov, ktorá používa nelineárnu metódu najmenších štvorcov pre fitovanie dat. Pre únavu je najvhodnejšie použiť model  $\sin 8$  a pre neúnavový stav model  $\sin 7$ . Ak dochádza k únave, tak sa zvyšuje počet sínusových výrazov, počet motorových jednotiek(MU), periodicita signálu a zvyšuje sa aj amplitúda signálu v porovnaní so signálom, kde sa únava nevyskytuje [15].

Vzťah medzi signálom EMG a diagnostikou stresu je zatiaľ nepredikovateľný, ale existujú nové prístupy na stanovenie stresu zo signálu EMG použitím neurónových sietí, ktoré sú aktuálne veľmi využívané pre spracovanie signálov, obrazov a v klasifikácii biologických obrazov, hlavne pre ich jednoduchosť a nepotrebnosť použitia komplikovaných matematických modelov [16]. Nasledujúce metódy spočívajú v analýze signálu EMG a použití umelých neurónových sietí v rozpoznávaní vzorov. V prvej z nich umelá neurónová sieť pozostáva z troch vrstiev a ako aktivačnú funkciu používa log-sigmoidu [17]. Nadväzujúca metóda zahŕňa aj výpočet koeficientov signálu pomocou autoregresného modelu a výsledok pozostáva v porovnaní štatistických údajov [18]. V ďalšej metóde sa používa DWT a dosahuje vysokú presnosť klasifikácie [19]. Iná metodika pozostáva z aplikácie algoritmu Multi-Layer Perceptron, pri ktorej bolo zistené,

že použitie umelých neurónových sietí a najmä algoritmu Back-propagation, na detekciu stresu zo signálu EMG dosahuje veľmi kvalitné výsledky [20]. Nastávajúca metóda používa na tréovanie sietí Hebbian a Perceptrón, ktoré boli použité v štúdií [16], kde boli snímané signály sEMG z kombinácie dvoch prstov, napríklad palec a ukazovák. Vytvorila sa dve série hodnôt, jedna séria pre jeden prst a druhá pre kombináciu dvoch prstov. Sú rozdelené do dvoch tried a to do triedy 1 pre jeden prst a triedy -1 pre dve prsty. Lepšie výsledky dosahuje Perceptrón, hlavne vďaka minimálnemu času, ktorý potrebuje na vytvorenie perceptronovej siete. Používa sa trojvrstvová Feed-Forward sieť a algoritmus Back-propagation.

### 1.3 Detekcia z EDA

Je preukázané, že elektrodermálna aktivita je jedným z hlavných identifikátorov stresu. Kombinované zmeny medzi elektrodermálnou rezistenciou a elektrodermálnym potenciálom sa nazývajú elektrodermálna aktivita EDA. Zo signálu EDA je možné extrahovať rôzne príznaky, ktoré sú relevantné pre vyhodnotenie stresu a z časti aj únavy. EDA je úzko spätá so sympatickým nervovým systémom, ktorý ovplyvňuje potné žľazy. Jeden z dvoch typov signálov EDA sa označuje ako exosomatický a meria sa ako impedancia kože spôsobená potením. Vodivosť kože má dve zložky a to tonickú(SCL) a fázičnú(SCR).

Zmena vodivosti kože v rade mikrosiemensoch je odpoveďou kože na prijaté vnútorné a vonkajšie podnety rôzneho charakteru. Jednou z možností je zistiť galvanický odpor kože meraním prúdu medzi dvomi elektródami umiestnenými v malej vzdialenosti od seba, alebo zistiť galvanický kožný potenciál ako napätie merané medzi dvomi elektródami. Prvá metóda preukázania stresu pozostáva z prevedenia FFT na signál EDA. Zo získaného spektra je možné zistiť, že v momente, kedy bol vybudený stres sa zvýšil rozptyl a počet vrcholov FFT v porovnaní s normálnym stavom a zároveň sa zväčšili aj hodnoty vrcholov v rozmedzí 0-0.05 Hz [21].

Jednou z ďalších možností detekcie stresu je využitie Machine learning prístupu. V prvom rade sa na základe preddefinovaného protokolu vyberú vhodné údaje na tréovanie modelu. Protokol obsahuje signály EDA, ktoré nevykazujú žiadne známky stresu, a tie, ktoré vykazujú stres. Signály, ktoré boli ovplyvnené pohybom a identifikované akcelometrom, boli následne vymazané pri vytvorení súboru dát. Vytvorenie modelu prebieha nasledovne. Najprv sa extrahujú tie príznaky, ktoré môžu vytvoriť priveľké odchýlky. Pomocou stromového klasifikátora sa získa dôležitosť každého príznaku ako váhový faktor. Tieto váhy sú následne porovnávané a vyberú sa len tie najvýznamnejšie pre tréovanie modelu. Ak je v tréovacím modeli zahrnutá široká škála stresových podmienok, je možné získať veľmi robustnú klasifikáciu. Pre tréovanie modelu sa použije klasifikátor K-nearest neighbor(KNN),  $k = 3$ , a vykoná sa

5-násobná krížová validácia. Klasifikátor KNN sa použije pre klasifikáciu stresového stavu a kľudového stavu [22].

Spomedzi všetkých biosignálov je EDA najcitlivejšia na hladinu stresu kvôli vysokej korelácii so sympatickým nervovým systémom a okrem toho senzor na meranie signálu EDA je veľmi jednoducho implementovateľný do inteligentného zariadenia. Najnovší verejne dostupný multimodálny súbor WESAD obsahujúci tri rôzne stavy – základný, stresový a stav pobavenia, sa považuje za štandard. WESAD obsahuje nasnímané signály EDA z hrudníka a zo zápästia a model detekcie stresu so 14 príznakmi. Nasledujúca metóda tento štandard modifikuje, najprv sa znižuje podiel prekrývajúcich sa segmentov s ohľadom na vplyv preučenia, následne extrahuje príznaky z časovej oblasti, z frekvenčnej oblasti, vlnkovej oblasti a z entropie. Z časovej oblasti sa vyberú príznaky ako je amplitúda, intervaly a doba trvania, z frekvenčnej oblasti sa vyberie PSD, a z vlnkovej oblasti, ktorá sa získa aplikáciou vlnkovej transformácie, získame vlnkové koeficienty. Na začiatku prebehne klasický postup predspracovania signálov. Potom prebehne dekompozícia signálu na tonickú a fázickú komponentu s detekciou vrcholu. V tejto metóde sa kombinuje stav základný so stavom pobavenia a je označený ako trieda bez stresu a porovnáva sa s triedou stresu. Jedná sa o binárnu klasifikáciu. Navrhne sa model pre výber najdôležitejších príznakov súvisiacich so stresom. Aplikuje sa učiaci model XGBoost ako klasifikačný model na predikciu triedy stresu, vďaka jeho schopnosti vysporiadania sa s vysoko dimenzionálnymi príznakmi a nevyváženými dátami. XGBoost je súbor techník založených na zvýšení gradientu, obsahujúci súbor slabších prediktorov. Pre výber dominantných príznakov sa použije Príznaková dôležitosť a Príznaková korelácia. Analýza dominantného príznaku poukazuje nato, že rozsah nízkej frekvencie a komplexita vysokej frekvencie signálu EDA sú užitočné pre detekciu stresu [23].

## 1.4 Detekcia z respirácie

Záznam respirácie patrí medzi indikátory stresu a únavy u vodiča, najmä frekvencia dýchania vykazuje vysoký stupeň účinnosti pri zisťovaní ospalosti a stresu. Meranie respirácie prebieha rôznymi spôsobmi, ale ukázalo sa, že nové postupy neposkytujú relevantné dáta k dosiahnutiu spoľahlivých výsledkov.

Medzi menej spoľahlivé metódy merania respirácie patrí bezkontaktné meranie. Samotné meranie spočíva v použití monitorovacej techniky, neinvazívnej kamery, ktorá je zabudovaná v interiéri vozidla. Dôležitú úlohu pri meraní zohráva umiestnenie kamery do vhodnej polohy. Predspracovanie pozostáva z eliminácie vonkajšieho verejného osvetlenia, svetla prichádzajúceho z okolitých áut, zlepšenia kontrastu, filtrovania a zo stabilizácie obrazu, pretože dochádza k vysokému pomeru signál/šum. Tento kamerový systém zaznamenáva a detekuje periodické pohyby hrudníka a brucha a na základe toho získa signál frekvencie dýchania. Získa sa len signál súvisiaci s dýchaním ako oblasť záujmu. Na odhad dychových pohybov sa používajú rozdiely medzi získanými signálmi

z videa. Meranie rýchlosti pohybu pomáha vykresliť semi-periodický signál, ktorý odpovedá dychovej frekvencii. Na odhad dychovej frekvencie sa aplikuje Krátkodobá Fourierová transformácia (STFT) a použije sa stredná hodnota dominantných frekvencií v časovom okne. Na spracovanie je použitý algoritmus optického toku, pričom samotný algoritmus je výpočtovo náročný. Pomerne kvalitné výsledky dosahujú aj infračervené kamery, ktorými sa analyzuje hĺbková mapa. Dosahujú až 90% koreláciu so záznamom získaným pletysmografickým pásom. V tejto metóde dochádza k vysokému pomeru signál/šum, čo vytvára obmedzenia tohto postupu, pretože meranie je ovplyvnené pohybmi vodiča, oblečením vodiča, antropometrickými parametrami a šumom z okolia, ktorými sú vnútorné osvetlenie auta, vonkajšie verejné osvetlenie a samotným umiestneným kamery [10].

Ďalšou metódou detekcie stresu pomocou merania respirácie je použitie hrudných/brušných pásov umiestnených okolo bránice vodiča, ktoré sú založené na pletysmografickom princípe. Signál z respirácie sa normalizuje odpočítaním priemeru signálu, ktorý bol nameraný v pokojnom stave subjektu. Vo frekvenčnej oblasti je vypočítané výkonové spektrum z určených bodov v strede každého segmentu, z ktorého sú následne vypočítané štyri príznaky, ktoré zastupujú energiu v každom zo štyroch pásem. Na signál sa aplikuje Hanningovo okno a na výpočet výkonového spektra sa používa modifikovaná metóda periodogramu pomocou zavedeného priemerného Welcha. Sumou energií v pásmach (0-0.1, 0.1-0.2, 0.2-0.3 a 0.3-0.4) Hz sú získané štyri príznaky hustoty výkonového pásma [11]. Iným spôsobom detekcie stresu je využitie dynamiky dýchania namiesto dychovej frekvencie. Na začiatku je nutné filtrovať signál filtrom pásmová zádrž medzi 0.03 a 0.08 Hz s Butterworthovým filtrom 3. rádu na elimináciu vysokofrekvenčných zložiek a základnej čiary. Kvôli zachovaniu morfológie sa urobí filtrácia forward-backward. Následne boli zo signálu získané informácie o konci inspirácie a exspirácie. V podstate na signál boli použité dve rôzne filtre, ktorými boli získané dve rôzne signály. Pre každý dýchací cyklus boli definované dve časové obdobia, inspirácia ( $T_{in}^j$ ) a exspirácia ( $T_{ex}^j$ ) a časový interval BB, ktorého jednu hodnotu tvorí čas, ktorý trvá od maxima exspirácie jedného dychu po maximum exspirácie nasledujúceho dychu.  $T_{in}^j$  je čas potrebný na prechod signálu z amplitúdy 0.1 A do 0.9A, a  $T_{ex}^j$  je čas potrebný na prechod signálu z amplitúdy 0.9 A na 0.1 A. Následne bol implementovaný algoritmus odmietnutia odľahlých hodnôt kvôli odstráneniu abnormálnych dychov. Vypočítané sú 3 amplitúdové rozdiely a zistí sa priemer a štandardná odchýlka predchádzajúcich 30 dychov pre všetky tri amplitúdy. Časy  $T_{in}^j$  a  $T_{ex}^j$  boli klasifikované ako funkcia BB. Vzťah medzi týmito parametrami sa javí ako lineárny a preto sa vytvorila dve úsečky, jedna pre  $T_{in}$  proti BB a druhá  $T_{ex}$  proti BB. Úsečky sú fitované cez „biquare“ fitovacia metódu kvôli zvýšeniu robustnosti proti odľahlým hodnotám. Potom sa získa sklon oboch línií  $\alpha_{T_{in}}$  a  $\alpha_{T_{ex}}$ . Takže  $\alpha_{T_{in}}$  predstavuje variáciu  $T_{in}^j$  v reakcii na zmenu hodnoty  $BB_j$ . Uhol  $\theta$ , tvorený priamkami so sklonmi  $\alpha_{T_{in}}$  a  $\alpha_{T_{ex}}$  sa považuje za meranie rozdielu medzi oboma sklonmi a je počítaný predpísanou rovnicou [12]. Je dokázané, že pri strese dochádza k zmene dynamiky dýchania. Táto metóda sa môže používať aj pre neinvazívne metódy [12].

## 2. SPRACOVANIE BIOLOGICKÝCH SIGNÁLOV

### 2.1 Spracovanie EKG

Spracovanie EKG signálu spočíva najmä v detekcii QRS komplexov, konkrétne detekcií R vln, vďaka ktorým vieme posúdiť spoľahlivosť a relevanciu nameraného signálu. Existujú rôzne prístupy úpravy a spracovania EKG signálu, ale ďalej uvedený prístup bol použitý v tejto práci.

Medzi základné úpravy patrí odstránenie driftu (jednosmerný prúd) a brumu (šum z elektrickej siete). Drift sa jednoducho odstráni filtrom horná priepusť, brum pomocou filtra úzkopásmová zádrž na 50 Hz. Signál je ďalej nutné predspracovať pásmovou priepusťou v hodnotách 10 Hz a 20 Hz pre zvýraznenie QRS komplexu a potlačenie ostatných zložiek signálu, ktorými sú vlny P a T a obmedzenie vplyvu myopotenciálov. Následne je signál upravený pomocou Hilbertovej transformácie a jeho absolútna hodnota je umocnená na druhú. Táto metóda umožňuje získanie obálky signálu, ktorá sa používa pri detekcii R vln, ktoré sú následne detekované pomocou prahovania [25] a nájdenia maximálnych hodnôt.

### 2.2 Spracovanie EMG

Biosignál EMG má širokospektrálne využitie, počínajúc práce s elektrickými biopotenciálmi až po kontrolu biorobotických mechanizmov. Pri akvizícií elektromyografických signálov je možné zachytiť šum, ktorý interferuje s informáciami, ktoré sú predmetom záujmu a z toho dôvodu je potrebné vytvoriť digitálne filtre na filtrovanie týchto rušivých zložiek.

Prvým krokom je eliminácia driftu. Signály EMG majú svoje najrelevantnejšie informácie vo frekvenčnom pásme medzi 4 Hz a 500 Hz a preto sa následne aplikuje na signál filter pásmová priepusť s medznými frekvenciami 4 Hz a 500 Hz. Existujú rôzne formy rušenia, najčastejšie sa vyskytuje šum 50 Hz z elektrickej siete (Slovenská a Česká republika, ...), ostatné formy šumu sú spôsobené pohybom káblov a ľudí, na ktorých sa signály merajú. Na odstránenie šumu spôsobeného elektrickou sieťou sa používa filter pásmová zádrž s medznými hodnotami 47 Hz a 53 Hz [34].

### 2.3 Spracovanie EDA

Signál EDA je zložený z dvoch častí, pomaly premennej zložky skin-conductance level (SCL) a rýchlo premennej zložky skin-conductance response (SCR), ktoré sú charakterizované ako úroveň vodivosti kože a zmena vodivosti kože.

Základom je rozdelenie signálu EDA pomocou dekonvolučného algoritmu na zložky SCL a SCR. Odhad nástupu a amplitúdy SCR je zvyčajne založený na analýze od minima po vrchol. Minima a maxima v údajoch sa považujú za latenciu nástupu a vrcholu

a amplitúda sa počíta ako rozdiel hodnôt pre čas nástupu a čas vrcholu. SCL je možné odhadnúť z adekvátnych časových intervaloch spriemerovaním skóre kožnej vodivosti v segmentoch [\[35\]](#).

## 2.4 Spracovanie signálu respirácie

Ľudské dýchanie predstavuje jedinečnú výzvu pre dekompozíciu signálu na jednotlivé zložky. Zo signálu je dôležité určiť jednotlivé parametre dýchania. Medzi parametre dýchania patrí extrémny nádychu a výdychu, trvanie nádychoch a výdychoch, maximálny respiračný tok, objemy jednotlivých nádychoch a výdychoch a pod.

Prvým krokom spracovania signálu respirácie pred dekompozíciou je odstránenie nameraného šumu a driftu signálu. Šum je odstránený pomocou vyhladzovacieho okna veľkosti 25 ms, ktorý prechádza signál zľava doprava. Veľkosť okna je dostatočná na odstránenie šumu a zároveň zachováva prirodzené tvary dýchania. V ďalšej fáze je odstránený globálny lineárny drift odpočítaním sklonu lineárnej regresie. Následne sa používajú rôzne algoritmy detekcie extrémov nádychu a výdychu, napríklad aplikácia posúvneho okna a hľadanie maxima/minima [\[36\]](#), alebo použitie prvej diferencií na nájdenie inflexných bodov [\[24\]](#).

### **3. VÝBER PRÍZNAKOV**

Pojem príznak je možné chápať ako informáciu, ktorá je prospešná pre predikciu výsledku v strojovom učení, alebo v iných metódach. Akákoľvek informácia o zadaných vstupných dátach, môže byť užitočná pre výslednú úspešnosť modelu. Každému súboru údajov prislúchajú špecifické príznaky, ktoré tieto dáta dobre charakterizujú. Na začiatok je nutné uskutočniť extrakciu relevantných a najlepšie nezávislých príznakov, ktoré prispejú ku lepšiemu výsledku pri spracovaní danej úlohy.

#### **3.1 Extrakcia príznakov**

Prvý krok vo výbere príznakov je porozumenie danej problematiky a zamyslením sa nad tým, či vybraný set príznakov je použiteľný v celom spektre dát. Ako príklad môže slúžiť signál EKG, z ktorého sa pomocou rôznych výpočtov extrahujú napríklad parametre MNN, SDNN a RMSSD, LF, HF, pomer LF/HF a veľa ďalších. Počet príznakov tvorí rovnaký počet dimenzií. Súbor vstupných dát príznakov s jednotlivými meraniami (pozorovaniami) môže dosahovať obrovské rozmery, niekoľko stoviek/tisíciek dimenzií. Nie všetky príznaky cez všetky merania ponúkajú relevantnú informáciu a preto môžu byť v procese selekcie vylúčené [41]. Ak techniky selekcie dobre fungujú, výsledkom je optimálny súbor údajov so všetkými dôležitými vlastnosťami, nižší výpočetný čas a lepšia vizualizácia dát. Tieto techniky sú opísané v podkapitole 3.4.

#### **3.2 Základná analýza príznakov**

Predtým, ako sa použije vstupná sada príznakov je potrebné urobiť základnú kontrolu. V prvom kroku sa overí, či všetky merania obsahujú všetky príznaky. Ak by došlo v 30% prípadoch k chýbajúcim dátam v meraniach, alebo v príznakoch, tak sa vymaže celé meranie, alebo daný príznak, pretože má malý prínos v novovznikajúcom modeli. Následne sa odstránia odľahlé hodnoty, ktoré zbytočne spôsobujú chyby merania. Merania a príznaky s veľmi malým rozptylom sú nadbytočné a tiež sú zo vstupnej sady odstránené. Poslednú zo základných úprav tvorí určenie závislosti medzi príznakmi. Ak majú niektoré príznaky vysoký stupeň závislosti, vypočítaný ako korelačný koeficient, je potrebné jeden z príznakov odstrániť, pretože je pravdepodobné, že obidve príznaky obsahujú rovnakú alebo veľmi podobnú informáciu [41]. Metriky merania závislosti sú popísané v nasledujúcej podkapitole.

#### **3.3 Meranie závislosti medzi príznakmi**

Medzi metriky merania závislosti patrí Pearsonov korelačný koeficient, Spearmanov korelačný koeficient, Korelácia vzdialenosti a Vzájomná informácia.



Najviac používaným typom korelačného koeficienta je Pearsonov, ktorý nadobúda hodnoty od -1 do +1 a reprezentuje lineárnu závislosť medzi dvomi premennými. Vypočíta sa podľa nasledujúceho vzorca 3.1, kde  $\mu_x$  a  $\mu_y$  znamenajú priemerné hodnoty premenných  $x$  a  $y$  a  $\sigma_x$  a  $\sigma_y$  znamenajú smerodajné odchýlky premenných  $x$  a  $y$  [41].

$$\rho = \frac{E[\{X - \mu_x\}\{Y - \mu_y\}]}{\sigma_x \sigma_y} \quad (3.1)$$

Spearmanov korelačný koeficient je založený na poradí premenných. Na rozdiel od Pearsonovho korelačného koeficientu je možné ho aplikovať aj na iné rozloženie dát ako je normálne a dokáže zachytiť nie len lineárny vzťah medzi premennými. Zároveň nie je citlivý na extrémne hodnoty. Ďalšou metrikou je Korelácia vzdialeností, ktorej hodnoty sú od 0 do 1 a meria závislosť medzi dvoma náhodnými vektormi  $X$  a  $Y$  pomocou Euklidovskej závislosti. Ak sú premenné nezávislé, tak hodnota je 0. S lineárnou alebo nelineárnou závislosťou hodnota stúpa. Rovnica 3.2 popisuje meranie korelácie vzdialeností, kde  $dCov(X, Y)$  značia kovarianciu dvoch premenných,  $Var(X)$  a  $Var(Y)$  rozptyl premennej  $X$  a  $Y$  [41].

$$dCor(X, Y) = \frac{dCov(X, Y)}{\sqrt{dVar(X)dVar(Y)}} \quad (3.2)$$

Poslednou korelačnou metrikou je Vzájomná informácia, ktorá má široké uplatnenie a meria množstvo informácií, ktoré je možné získať z jednej náhodnej premennej danej inej [41].

Na meranie závislosti príznakov sa môže použiť aj chi-kvadrát test. Pomocou jednotlivých chí-kvadrát testov sa preskúma, či je každá premenná prediktora nezávislá na premennej odozvy a potom sa príznaky zoradia pomocou štatistických p-hodnôt chí-kvadrát testu. Malá p-hodnota znamená, že zodpovedajúca predikčná premenná závisí od premennej odozvy a preto je dôležitou vlastnosťou. Rovnica 3.3 ukazuje rovnicu chí-kvadrát testu, kde  $c$  je stupeň voľnosti,  $O$  sú pozorované hodnoty a  $E$  sú očakávané hodnoty [42].

$$\chi_c^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3.3)$$

### 3.4 Selekcia príznačkov

Selekcii príznačkov je možné rozdeliť na dve skupiny. Jednu skupinu tvorí selekcii príznačkov, ktorá zanecháva originálne príznačky a druhú skupinu tvorí selekcii príznačkov bez zachovania pôvodných príznačkov.

Najznámejšou metódou, pri ktorej dochádza k strate pôvodných príznačkov je metóda Analýzy hlavných komponent, teda PCA, ktorá hľadá takú množinu lineárnych kombinácií pôvodných premenných (pozorovaní), ktorá zachováva čo najväčšie množstvo informácií o pôvodných premenných a zároveň jej dimenzia je menšia. Transformuje prvky množiny premenných na prvky takej množiny hodnôt, ktoré sú lineárne nekorelované a označujú sa ako hlavné komponenty. Medzi ďalšie metódy patrí Lineárna diskriminačná analýza LDA, Stochastické vkladanie susedov t-SNE, a ďalšie.

Do skupiny metód, ktoré zachovávajú pôvodné príznačky patrí klasifikátor Náhodný les, Dopredná selekcii, Spätná selekcii, algoritmus MRMR. Klasifikátor Náhodný les je jednoduchá, relatívne presná a robustná metóda, ktorá spočíva v zostrojení množstva rozhodovacích stromov pri tréovaní na základe baggingu, teda náhodného rozdelenia dát na InBag (tréovacie vzorky) a OutOfBag (testovacie vzorky). Táto metóda znižuje rozptyl.

Ďalšou metódou je Dopredná selekcii, u ktorej sa najprv zoberú jednotlivé príznačky, z ktorých každý sa použije zvlášť na klasifikáciu. Potom sa zisťuje, ktorý dosiahol najlepšie výsledky a postupne sa k nemu pridávajú ďalšie príznačky a klasifikátor sa opäť tréuje.

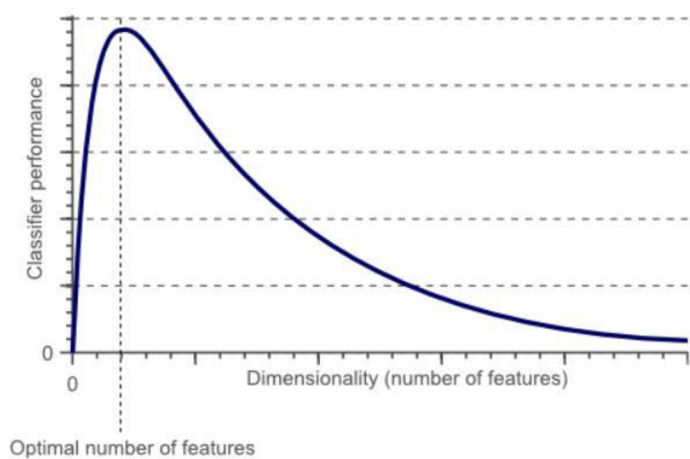
Spätná selekcii funguje podobne, ale najprv sa uskutoční klasifikácia na celej sade príznačkov a z nich sa postupne odoberajú príznačky a znova prebieha tréovanie klasifikátora.

Algoritmus Maximálnej-relevancie a Minimálnej-Nadbytočnosti, MRMR, je založený na tom, že dobrý príznaček by mal mať maximálnu relevanciu k danej triede a zároveň minimálnu nadbytočnosť k už vybraným príznačkom. Algoritmus pracuje iteratívne a využíva vzájomnú informáciu. Najprv sa vypočíta, ktorý príznaček má najväčšiu relevanciu k triede a následne sa k nemu pridávajú príznačky tak, aby bolo maximalizované kritérium  $\max(D(S) - R(S))$ , kde  $D(S)$  je stredná hodnota vybraných vzájomných informácií a  $R(S)$  je vzájomná závislosť medzi dvomi premennými [41].

Algoritmus NCA, teda analýza susedských komponentov je kontrolovaná učebná metóda na klasifikáciu údajov s viac premennými do rôznych tried podľa danej metriky vzdialenosti. Má podobné uplatnenie ako algoritmus K-najbližších susedov a využíva koncept zvaný stochastickí najbližší susedia. Cieľom tejto metódy je učenie vzdialenostnej metriky nájdením lineárnej transformácie vstupných údajov tak, aby sa v transformovanom priestore maximalizoval priemerný výkon klasifikácie [54].

Hughesov fenomén hovorí o tom, že s rastúcim počtom dimenzií klesá výkon modelu. Napríklad 10 dimenzií s 1000 pozorovaniami bude mať výkon väčší, ako 100

dimenzií s 1000 pozorovaniami. Obrázok 3.1 ukazuje Hughesov fenomén, kde priamka pretínajúca x osu značí najoptimálnejší počet vybraných príznakov, dimenzií.



Obrázok 3. 1: Hughesov fenomén. Prevzaté z [\[41\]](#).

## 4. KLASIFIKAČNÉ METÓDY STROJOVÉHO UČENIA

Klasifikácia je široko používaná metóda strojového učenia, ktorá sa používa na predikciu tried. Patrí medzi metódy učenia s učiteľom, ktorej algoritmus pracuje v 2 krokoch. Najprv sa klasifikátor naučí na dostupných tréningových dátach a vytvorí sa naučený model, ktorý sa použije na klasifikáciu nových pozorovaní. Tréňované dáta sa použijú na určenie výslednej triedy na základe okrajových podmienok.

Klasifikátory je možné rozdeliť podľa počtu tried na dva typy, binárne a mnohotriedne. Binárne klasifikátory pracujú len s dvomi triedami, napr. pozitívny a negatívny pacient. Mnohotriedne klasifikátory rozdeľujú pozorované dáta do troch alebo viacerých tried. Existuje varianta tzv. „multi-label“ klasifikácie, u ktorej je možné priradiť každej inštancii viac lablov. Jedná sa o zovšeobecnenú mnohotriednu klasifikáciu, kde vzniká problém označovania jednotlivých inštancií do jednej z viac ako troch tried, pričom neexistuje žiadne obmedzenie, do koľkých tried je možné inštanciu priradiť [31].

Existuje veľké množstvo klasifikačných algoritmov a v nasledujúcich podkapitolách budú priblížené najznámejšie typy klasifikačných metód.

### 4.1 Rozhodovacie stromy

Princípom je hierarchický, viacstupňový binárny rozhodovací systém, v ktorom sa postupne vyhodnocuje splnenie, alebo nespĺnenie rozhodovacích kritérií alebo podmienok, pokiaľ sa nedostaneme k akceptovanej triede alebo riešeniu.

V procese rozhodovania sa postupuje od koreňa až po jednotlivé uzly stromu, ktoré tvoria vetvy stromu s listami. Príznak, alebo kritérium, ktorý má najväčšiu váhu a najlepšie popisuje súbor dát sa stáva koreňom stromu a separuje vstupné dáta do 2 binárnych tried (áno/nie). Ďalšie uzly sú postupne tvorené zostávajúcimi kritériami s menšou váhou, pričom opäť každý uzol vytvára dvoch binárnych potomkov. V procese rozhodovania sa postupuje v smere od koreňa až k terminálnemu listu [29].

Pri výbere prediktora algoritmus zohľadňuje všetky príznaky a vyberie ten, ktorý zabezpečí najväčšiu presnosť a rekurzívne sa proces opakuje, kým sa úspešne nerozdelia dáta do všetkých listov, alebo sa nedosiahne maximálna hĺbka stromu. Rozhodovacie stromy sú rýchlo pochopiteľné a vizualizovateľné. Dokážu pracovať s číselnými aj kategóriálnymi údajmi. Jednou z nevýhod je, že hrozí u nich preučenie [31]. Rozhodovacie stromy nachádzajú svoje uplatnenie rovnako v úlohách klasifikácie ako aj regresie.

## 4.2 Naivný Bayesov klasifikátor

Jedným z prvých populárnych algoritmov pre klasifikáciu v strojovom učení bol Naivný Bayesov klasifikátor, založený na Bayesovskej vete s predpokladom nezávislosti medzi prediktormi. Bayesovská veta hovorí o možnosti vykonať rozhodnutie na základe priorných znalostí. Inými slovami, tento klasifikátor predpokladá, že prítomnosť konkrétneho znaku v triede nesúvisí s prítomnosťou iného znaku, alebo, že všetky tieto vlastnosti majú nezávislý príspevok k pravdepodobnosti. Je pomerne jednoduchý na zostavenie a je veľmi účinný pre veľké súbory dát, pretože je vysoko škálovateľný a zároveň vyžaduje malé množstvo tréningových dát. Naivný Bayesov klasifikátor, aj napriek svojej jednoduchosti, prekonáva vysoko sofistikované klasifikačné metódy.

Prvým krokom algoritmu je výpočet priornej pravdepodobnosti pre dané laby tried. Následne sa zistí likelihood pravdepodobnosť s každým atribútom pre každú triedu. Získané hodnoty sa vložia do Bayesovského vzorca a vypočíta sa posteriorna pravdepodobnosť a z nej sa zistí, ktorá trieda má vyššiu pravdepodobnosť, pretože je dané, že vstup patrí do triedy vyššej pravdepodobnosti [\[31\]](#).

## 4.3 K – najbližších susedov

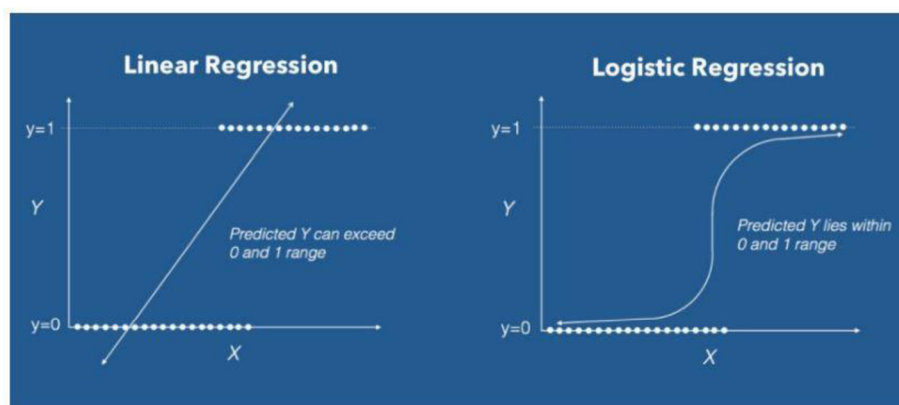
kNN je jednoduchý klasifikačný algoritmus, ktorý ukladá všetky dostupné prípady a klasifikuje nové prípady na základe miery podobnosti, napr. vzdialenostnej metriky, ako je Euklidová alebo Manhattanská vzdialenosť. Využíva malú vzdialenosť ako zástupný znak pre rovnakosť.

Algoritmus vezme množstvo označených bodov a použije ich na naučenie modelu, ktorý použije na označenie ďalších nových bodov. V prípade značenia nového bodu sa pozrie na označené body v modeli, ktoré sú najbližšie k novému bodu. Akonáhle skontroluje počet „k“ najbližších susedov, priradí mu label podľa toho, aký label má najviac susedov [\[31\]](#).

## 4.4 Logistická regresia

Je to štatistická metóda na analýzu súboru údajov, v ktorom existuje jedna alebo viac nezávislých premenných, ktoré určujú výsledok. U tohto klasifikátoru sú pravdepodobnosti, ktoré popisujú očakávané výsledky, modelované pomocou logistickej funkcie. Zamiera sa na binárnu klasifikáciu, avšak pri mnohotriednej klasifikácii je možnú ju rozšíriť na multinomiálnu a ordinálnu logistickú regresiu.

Cieľom logistickej regresie je nájsť najvhodnejší model na popis vzťahu medzi dichotomickou charakteristikou, ktorá nás zaujíma (závislá premenná = premenná odpovede alebo výsledku) a súborom nezávislých (predikčných alebo vysvetľujúcich) premenných [\[32\]](#). Rozdiel medzi lineárnou a logistickou regresiou je naznačená na obrázku 4.1., kde pri lineárnej regresii môžu hodnoty nadobúdať rozsah hodnôt presahujúcich 0 a 1, a pri logistickej regresii ležia hodnoty v intervale 0-1.

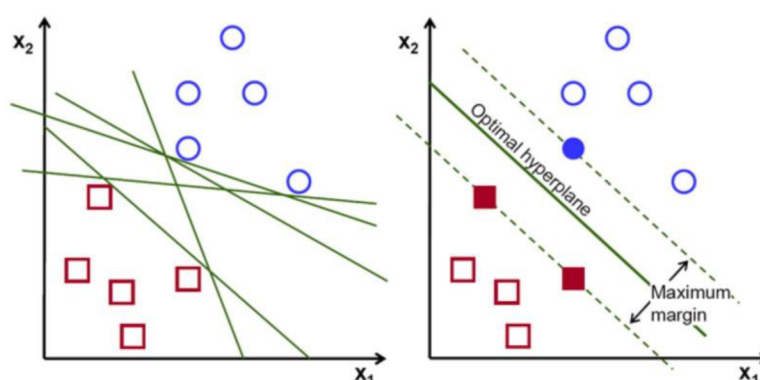


Obrázok 4. 1: Lineárna regresia vľavo, logistická regresia vpravo. Prevzaté z [31].

## 4.5 SVM – metóda podporných vektorov

Metóda podporných vektorov je jednoduchý algoritmus, ktorý produkuje významnú presnosť s menším výpočtovým výkonom a je implementovateľný na klasifikačné aj regresné úlohy.

Cieľom algoritmu SVM je nájsť nadrovinu v N-dimenzionálnom priestore (N-počet príznakov), ktorá zreteľne klasifikuje údajové body. Ak sa body rozdelia do dvoch tried, vytvorí sa rada nadrovin. Základom je nájsť takú nadrovinu, ktorá vytvára maximálnu vzdialenosť medzi okrajovými dátovými bodmi oboch tried. Podporné vektory sú dátové body, ktoré sú najbližšie k nadrovine a ovplyvňujú polohu a orientáciu nadroviny. Pomocou týchto podporných vektorov maximalizujeme okraj klasifikátora [33]. Na obrázku 4.2 vľavo sú zelenými priamkami zvýraznené rôzne nadroviny a na obrázku vpravo je znázornená optimálna nadrovina zelenou priamkou a vyplnené kocky a kruhy sú podporné vektory dvoch tried.



Obrázok 4. 2: Typy nadrovin označené zelenou priamkou (vľavo), optimálna nadrovina (vpravo), prevzaté z [33].

## 5. VYHODNOTENIE KLASIFIKÁTORA

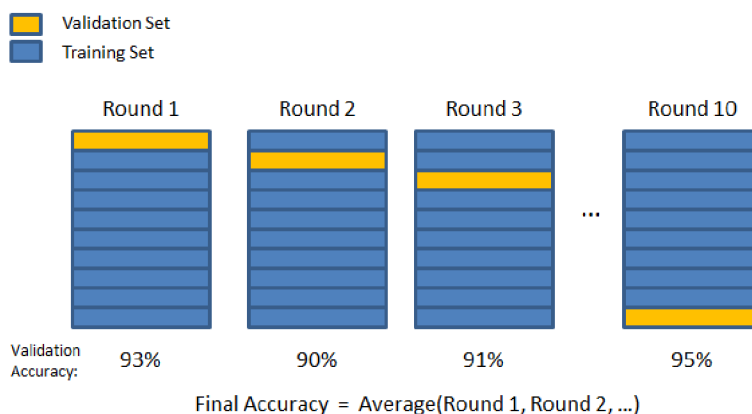
Pri vytváraní modelu strojového učenia je dôležité uskutočniť hodnotenie ako dobre daný model funguje na tréningových dátach. To sa určuje zvyčajne kombináciou dvoch možných prístupov, hodnotenia výkonu klasifikátora a kritériálnou funkciou na hodnotenie výkonu klasifikátora. Jedným z hlavných dôvodov hodnotenia klasifikátora je zistenie, či daný model poskytuje spoľahlivý výsledok na nových dátach. Ďalším dôvodom je porovnanie dvoch a viacerých klasifikátorov a zistenie, ktorý poskytuje najpresnejšie výsledky, pretože pri rovnakých dátach môže každý klasifikátor poskytovať rôzne výsledky [37].

Naučený model musí byť testovaný na testovacej sade dát kvôli hodnoteniu výkonu. Ak by sa model testoval výlučne na tréningovej sade dát, výsledok by bol mimoriadne optimistický. Jednou z možností je rozdeliť množinu dát na tréningovú a testovaciu, čím sa zaoberá metóda v podkapitole 5.1 [38]. Medzi prístupy hodnotenia modelu patrí klasifikačná presnosť, matica neurčitosti, ROC krivka, atď.

### 5.1 K-násobná krížová validácia

K-násobná krížová validácia je metóda zisťovania, ako dobre bude model ovplyvňovať nezávislé vzorky. Základom krížovej validácie je rozdelenie vstupného súboru dát na K množín, ktoré sú rovnako veľké a majú rovnakú distribúciu tried. Tie sú ďalej rozdelené na tréningovú množinu, ktorá pozostáva z tréningovej a validačnej množiny a zvyšné tvoria testovacie množiny dát. Klasifikátor natrénuje model na tréningovej množine a pomocou testovacej množiny testuje presnosť a výkonnosť modelu. Tento proces sa K-krát zopakuje, ale stále s inými množinami, ktoré tvoria tréningovú a testovaciu časť, vid' obrázok 5.1.

Špeciálny prípad krížovej validácie je tzv. „leave-one-out“ krížová validácia, kde počet K je rovný počtu N vzoriek v súbore dát. Proces sa opakuje N-krát, pričom N-1 vzoriek sa používa na tréningovanie a zostávajúca vzorka sa používa na testovanie [38].



Obrázok 5. 1: Príklad K-násobnej krížovej validácie a výpočet presnosti, prevzaté z [39].

## 5.2 Klasifikačná presnosť, miera chyby

Klasifikačná presnosť je jednou z najzákladnejších a najintuitívnejších metód na hodnotenie modelu a značí percento správne predikovaných hodnôt spomedzi všetkých predikovaných hodnôt. Označuje sa  $Acc$  a výpočet je definovaný v rovnici 5.1.

$$Acc = \frac{\text{správne predikované hodnoty}}{\text{všetky predikované hodnoty}} \cdot 100 \% \text{ (5.1)}$$

Miera chyby charakterizuje percento nesprávne predikovaných hodnôt spomedzi všetkých predikovaných hodnôt. Označuje sa  $Err$  a vypočíta sa ako odčítanie klasifikačnej presnosti od 1 [38].

## 5.3 Matica neurčitosti

Pred definovaním matice neurčitosti je potrebné oboznámiť sa s parametrami, ktoré sú potrebné pre jej výpočet. TP značí skutočne pozitívnu klasifikáciu dát, čiže model predikuje 1 a skutočná hodnota je 1. TN skutočne negatívnu klasifikáciu, model definuje predikuje 0 a skutočná je tiež 0. FP falošne pozitívnu klasifikáciu, model predikuje hodnotu 1, a skutočná hodnota je 0. FN značí falošne negatívnu klasifikáciu, takže model predikoval 0, ale skutočná je 1.

Matica neurčitosti, ktorá je známa aj ako matica chýb, znázorňuje 4 možné výsledky klasifikátora. Každý riadok predstavuje inštalácie v skutočnej triede, zatiaľ čo každý stĺpec predstavuje inštalácie v predikovanej triede, vid' obrázok 5.2. Táto matica umožňuje odvodiť veľa výkonnostných metrík. Súčtom hodnôt na hlavnej diagonále v matici neurčitosti a následným podelením celkovým počtom vzoriek dát je možné získať požadovanú klasifikačnú presnosť [37].

		Skutočná trieda	
		Pozitívne (1)	Negatívne (0)
Predikovaná trieda	Pozitívne (1)	TP	FP
	Negatívne (1)	FN	TN

Obrázok 5. 2: Matica neurčitosti so zvýraznenými správne klasifikovanými hodnotami.



## 5.4 ROC krivka

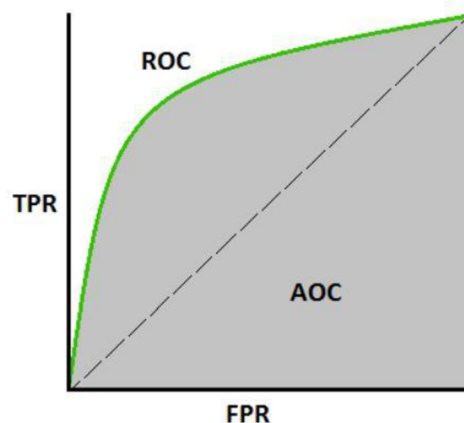
Jedným z kvalitných ukazovateľov výkonnosti navrhnutého klasifikačného modelu je tzv. krivka prevádzkovej charakteristiky, ROC krivka, ktorá vytvára vizuálnu podobu výkonu klasifikačného modelu. ROC je krivka pravdepodobnosti a je vynesená do grafu, kde x osu tvorí FPR a y osu TPR. TPR charakterizuje počet skutočne pozitívnych hodnôt, senzitivitu. Jej výpočet sa nachádza v rovnici 5.2.

$$TPR = \frac{TP}{TP+FN} \quad (5.2)$$

Pre výpočet FPR, ktorý charakterizuje počet falošne pozitívnych hodnôt, potrebujeme najprv zistiť špecificitu TNR, ktorej rovnica je uvedená v 5.3. Následne sa FPR vypočíta ako rozdiel jednotky a špecificity.

$$TNR = \frac{TN}{TN+FP} \quad (5.3)$$

Krivka ROC sa niekedy označuje aj AUC-ROC krivka, teda oblasť pod krivkou prevádzkovej charakteristiky. Čím je krivka bližšie k ľavému hornému rohu, tak tým je klasifikátor kvalitnejší [38]. AUC predstavuje stupeň alebo mieru separovateľnosti a je definovaná ako oblasť pod krivkou, vid' obrázok 5.3. Udáva, do akej miery je model schopný rozlišovať medzi triedami, čiže hľadáme najvyššiu hodnotu AUC. Plocha pod krivkou je rovná pravdepodobnosti, že model priradí náhodnému pozitívnemu príkladu väčšiu váhu ako náhodnému negatívnemu príkladu. AUC má rozsah hodnôt 0-1, pričom AUC = 0 znamená, že klasifikátor zaradil všetky hodnoty zle a ak AUC = 1, tak klasifikátor zaradil všetky hodnoty správne [40].



Obrázok 5. 3 - ROC krivka, sivá oblasť značí AUC, prevzaté z [38].

## 5.5 F1 skóre

F1 skóre je harmonický priemer medzi pozitívne predikovanou hodnotou PPV a senzitivitou. PPV je definovaná ako pomer medzi počtom správne pozitívne klasifikovanými vzorkami a celkovým počtom pozitívne predikovanými vzorkami. Harmonický priemer dvoch čísel, PPV a senzitivity, je bližšie ku menšiemu z nich. Z toho plynie, že vysoká hodnota F1 skóre zašľuje, že PPV aj senzitivita sú primerane vysoké a model je lepší [37]. Rozsah F1 skóre je v rozmedzí od 0 do 1. Výpočet F1 skóre prebieha podľa rovnice 5.4 [38].

$$F1 = \frac{2}{\left(\frac{1}{PPV}\right) + \left(\frac{1}{Senzitivita}\right)} \quad (5.4)$$

## 5.6 Výkon regresie

Regresné prediktívne modelovanie je úlohou aproximácie mapovacej funkcie ( $f$ ) zo vstupných premenných ( $X$ ) na spojitú výstupnú premennú ( $Y$ ). Spojitá výstupná premenná je skutočná hodnota, napríklad celé číslo. Na klasifikáciu sa používa logistická regresia a výstupný klasifikátor sa hodnotí pomocou nasledujúcich metód [43].

Medzi základné metódy patrí MAE, stredná absolútna chyba, ktorá počíta zostatok pre každý dátový bod, pričom berie do úvahy len absolútne hodnoty. V nasledujúcej rovnici,  $n$  znamená počet bodov,  $y$  znamená aktuálnu výstupnú hodnotu a  $\hat{y}$  predikovanú výstupnú hodnotu.

$$MAE = \frac{1}{n} \sum |y - \hat{y}| \quad (5.5)$$

Ďalšou metrikou je MSE, stredná štvorcová chyba, ktorá meria sumu priemerov štvorcových rozdielov medzi odhadovanou hodnotou a skutočnou hodnotou, viď rovnica 5.6.

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2 \quad (5.6)$$

Stredná absolútna percentuálna chyba, MAPE, je percentuálny ekvivalent MAE, s tým, že chyba je relatívna k danej výstupnej hodnote. Poslednou známou metrikou je koeficient determinácie  $R^2$ , ktorý sa používa na porovnanie dvoch modelov a kvantifikuje kvalitu regresie a je možné ju chápať ako rozptyl modelu vzhľadom na celkový rozptyl dát. Všetky hodnotenia výkonu regresie boli prevzaté z [38].

## 6. OPTIMALIZÁCIA HYPERPARAMETROV

Koncepčne je vyladenie hyperparametrov iba optimalizačnou slučkou na začiatku tréovania modelu strojového učenia, pomocou ktorej sa dá nájsť sada hyperparametrov vedúca k najnižšej chybe v testovacej sade. Počas optimalizácie hyperparametrov je základom nájsť takú kombináciu ich hodnôt, ktorá pomôže nájsť požadované minimum (chybu) alebo maximum (presnosť) funkcie. To zohráva dôležitú úlohu pri porovnávaní výkonnosti rôznych modelov strojového učenia pri rovnakej sade vstupných dát.

Modely strojového učenia pozostávajú z 2 typov parametrov. Jednu podoblasť tvoria hyperparametre, ktoré môžu byť nastavené používateľom pred začatím tréovania, a používa sa napríklad pri algoritme Náhodný les. Druhý typom sú parametre modelu, ktoré sa učia počas tréovania modelu a používajú sa pre logistickú regresiu a neurónové siete [44]. V nasledujúcich podkapitolách budú vysvetlené niektoré metódy optimalizácie hyperparametrov.

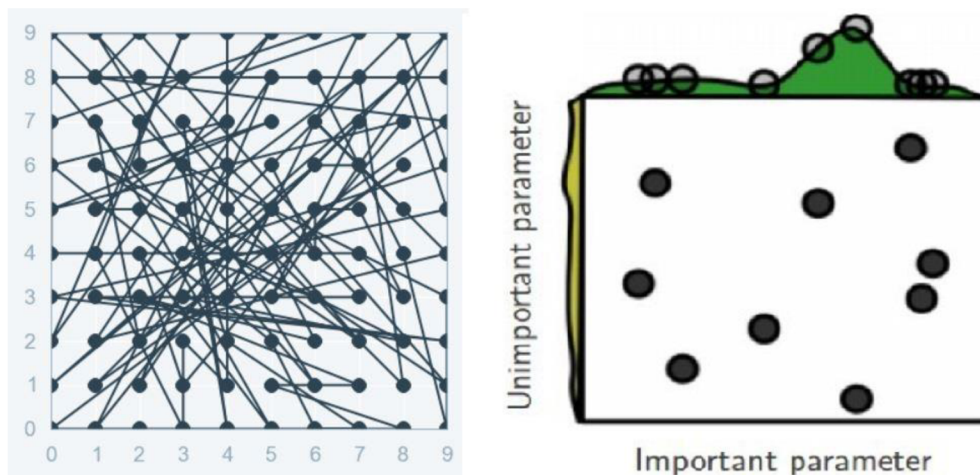
### 6.1 Manuálne prehľadávanie

Pri manuálnom prehľadávaní užívateľ vyberá niektoré modelové hyperparametre na základe skúseností a vlastného úsudku. Jedná sa o subjektívny výber parametrov. Vybrané parametre sa nastavujú pri tréovaní modelu a následne sa vyhodnotí jeho presnosť a proces sa začne odznova, kým užívateľ nedosiahne požadovanú presnosť. Tento postup je časovo náročný a nesmeruje k najoptimálnejšiemu riešeniu [44].

### 6.2 Náhodné prehľadávanie

Náhodné prehľadávanie je technika, pri ktorej sa na nájdenie najlepšieho riešenia pre zostavený model používajú náhodné kombinácie hyperparametrov, vid' obrázok 6.1 vľavo, kde je znázornený náhodne prehľadávaný priestor. Je to podobné ako pri vyhľadávaní v mriežke, vid' podkapitola 6.3, ale niekedy prináša relatívne lepšie výsledky. Pretože výber parametrov je úplne náhodný a na vzorkovanie týchto kombinácií sa nepoužíva inteligencia, šťastie pri tejto metóde zohráva veľkú rolu [45]. Algoritmus je nekonečný a preto je nutné mu zadať časový interval prehľadávania [44].

Jednou z výhod náhodného prehľadávania je, že ak dva hyperparametre navzájom málo súvisia, náhodne prehľadávanie umožňuje nájsť presnejšie optimá každého hyperparametru, čo je možné vidieť na obrázku 6.1 vpravo, kde na x ose sú dôležité parametre a na y ose sú menej významné parametre a nájdené optimá tvoria kolieska v zelenej časti. Jeho nevýhodou je, že neexistuje záruka nájdenia miestneho minima s určitou presnosťou, okrem prípadov, kedy je priestor na vyhľadávanie dôkladne vzorkovaný. Ak je tréovanie modelu za použitia veľkého množstva výpočtových údajov náročné, je tento algoritmus neefektívny [46].

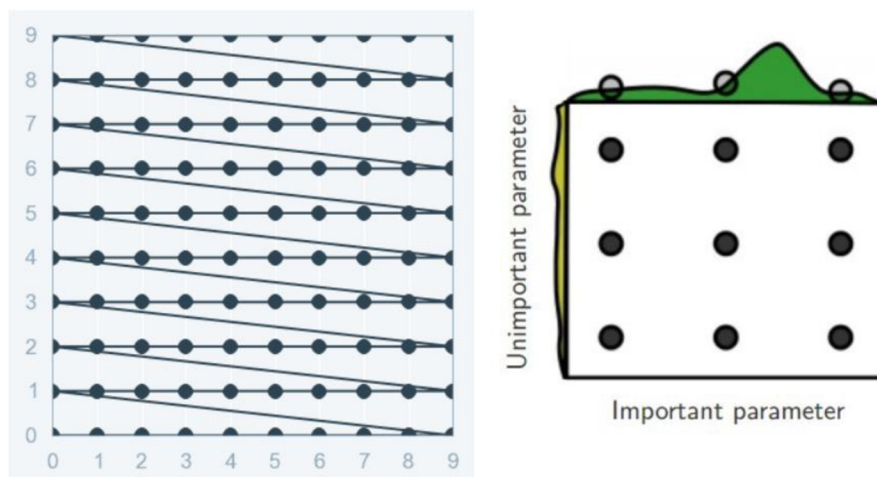


Obrázok 6. 1 : Náhodné prehľadávaný priestor vľavo, nájdené optimálnych hyperparametrov vpravo, prevzaté z [45].

### 6.3 Vyhľadávacia mriežka

Pri tomto prehľadávaní sa vytvorí mriežka, v ktorej je priestor hyperparametrov diskretizovaný a celkový vyhľadávací priestor je diskretizovaný kartézskymi súradnicami [46], viď obrázok 6.2 vľavo. Na obrázku 6.2 vpravo je znázornená mriežka s nájdenými optimálnymi hyperparametrami vyznačenými kolieskom v zelenej časti. Vzniknutý model sa natrénuje a otestuje na každej z možných kombinácií a zvolí tú najlepšiu. Na zvolenie parametrov, ktoré sa majú použiť pri vyhľadávaní v mriežke je možné sa pozrieť na parametre, ktoré najlepšie fungovali pri náhodnom vyhľadávaní a na ich základe zostaviť mriežku.

Vyhľadávanie v mriežke je v porovnaní s náhodným prehľadávaním pomalšie, alebo môže byť efektívnejšie, lebo prechádza celým priestorom. Výhodou je jednoduchosť, môže bežať paralelne a nepotrebuje nijakú formu ladenia. Na druhej strane má podobný problém ako náhodné prehľadávanie, neexistuje záruka nájdenia miestneho minima s určitou presnosťou, okrem prípadov, kedy je priestor na vyhľadávanie dôkladne vzorkovaný [46].



Obrázok 6. 2 : Vyhľadávanie v mriežke vľavo, nájdené optimálnych hyperparametrov vpravo, prevzaté z [44].

## 6.4 Bayesovská optimalizácia

Bayesovská optimalizácia využíva pravdepodobnosť na nájdenie minima funkcie. Konečným cieľom je nájdenie vstupnej hodnoty funkcie, ktorá dá najnižšiu možnú výstupnú hodnotu. Pri bayesovskej optimalizácii sa snažíme zostrojiť nejaký pravdepodobnostný model pre účelovú funkciu, teda považujeme ju za náhodnú. Pravdepodobnostný model najskôr odhadneme apriórnym rozdelením. V každom kroku optimalizácie na základe vypočítaných hodnôt funkcie  $f$  apriórne rozdelenie aktualizujeme a získame aposteriórne rozdelenie, teda hľadaný model. Tento postup je totožný s bayesovskými metódami pre odhad parametra na základe pozorovaných dát, ale namiesto odhadu parametra odhadujeme funkciu  $f$  [45].

Niektoré publikácie uvádzajú, že je účinnejšia ako náhodné, mriežkové alebo manuálne prehľadávanie. Bayesovská optimalizácia môže viesť k lepšiemu výkonu v tréningovej fáze a skráteniu času potrebného na optimalizáciu, keďže znižuje počet iterácií vyhľadávania pomocou zapamätania si posledných výsledkov [46].

## 6.5 Genetické algoritmy

U týchto algoritmov dochádza k aplikovaniu mechanizmov prirodzeného výberu na metódy strojového učenia. Algoritmy sú inšpirované Darwinovým procesom prírodného výberu a preto sa označujú aj evolučné algoritmy.

Princíp spočíva vo vytvorení populácie  $N$  tréningových modelov pomocou hyperparametrov, vypočíta sa ich presnosť a vyberú sa najlepšie. Vytvorí sa niekoľko potomkovia, ktorí majú podobné hyperparametre ako najlepšie modely, na získanie populácie ďalších  $N$  modelov. Potom sa vypočíta presnosť každého modelu a cyklus sa opakuje pre definovaný počet generácií a na konci sa prežijú len najlepšie modely [46].

## 7. NAMERANÉ DÁTA Z JÁZD

### 7.1 Databáza

V tejto diplomovej práci boli použité dáta, ktoré boli namerané v spolupráci s Ústavom súdneho inženýrství v Brne. K týmto meraniam bolo nasadené vozidlo značky BMW 5. rady s automatickou prevodovkou.

Dáta boli merané za prítomnosti rôznych subjektov, vodičov, pričom každý vodič absolvoval dennú a nočnú jazdu v ten istý deň. Jazdy boli uskutočnené v mesiacoch máj, jún a júl v roku 2019 a následne júl a september roku 2020. Vodiči absolvovali cestu zo stanovišťa pri kasárňach v Dědicích(časť Výškova), prešli cez mesto Vyškov a odbočili na výjazd smer Ivanovice na Hané, kde použili výjazd na diaľnicu(privádzač D46) a pokračovali smerom na Kroměříž. Na zjazde Kroměříž D1, 258 opustili diaľnicu a vzápätí sa na ňu vrátili v smere na Brno. Následne použili zjazd na Ivanovice na Hané a opäť pokračovali do Výškova na východzí bod pri kasárňach po ceste prvej triedy č. 47. Kvôli eliminácii vedľajších aspektov sa každá jazda vykonala za priaznivých podmienok, denná jazda za slnečného počasia. Zakaždým bol subjekt navigovaný rovnakou trasou. Trasa zahŕňala jazdu po meste, mimo mesta a aj diaľničné úseky. V meste musel vodič prejsť rôznymi križovatkami, kruhovými objazdmi, prechodmi z vedľajšej cesty na hlavnú cestu, rôznymi dopravnými obmedzeniami a práce na ceste. Diaľničné úseky boli väčšinou sprístupnené za obmedzených podmienok, a to hlavne za zníženej rýchlosti, za zníženia jazdných pruhov alebo nadmerného zúženia jazdného pruhu.

Na zachytenie relevantných dát u vodičov bolo použité zariadenie Biosignalsplux. Toto zariadenie bolo vytvorené na meranie biosignálov, ktorými sú EKG, EMG, EDA, EEG, elektrookulografia EOG, respirácia a ďalších typov signálov ako napríklad akcelerácia a teplota. Toto zariadenie meria súčasne viac typov biosignálov v reálnom čase pomocou elektród, ktoré sa aplikujú na pokožku subjektu a akvizované dáta posiela pomocou Bluetooth do počítača [24].

Počas jazdy bolo u subjektov merané z rady biosignálov EKG, EMG a EDA pomocou systému Biosignalsplux. Tieto signály boli vzorkované so vzorkovacou frekvenciou 1000 Hz. EKG bolo merané pomocou trojelektrodového systému z hrudnej časti subjektov, v ľavej hornej časti, aby mal bezpečnostný pás čo najmenší vplyv na meranie. EMG meranie prebiehalo pomocou dvojelektrodového systému z pravej nohy subjektov, s tým, že elektródy boli umiestnené nad sebou na svalu anterior tibialis, aby bola dosiahnutá najväčšia odozva kontrakcie svalu. EDA podobne pomocou dvojelektrodového systému z vnútornej strany ruky, na konci predlaktia smerom k zápästiu.

Súčasťou merania bol eyetracker, ktorý sleduje pohyb oboch očí, vid' Obrázok 7.1. Jednalo sa o zariadenie od firmy Ergoneers a konkrétne zariadenie nesie názov Dikablis Eye Tracking Glasses [30]. Toto zariadenie pripomínajúce okuliare, mal subjekt

počas jazdy na hlave. Systém mal v sebe integrovanú aj kameru, ktorá sledovala oblasť pred vodičom, respektíve palubovú dosku s volantom a cestu pred automobilom. Vo vozidle boli nainštalované značky s podsvietením, ktoré pripomínali QR kódy, vid' Obrázok 7.2. Tieto značky slúžili na určenie pohľadu šoféra, kde výsledkom je červený terč na Obrázku 7.2. Jednotlivé značky boli rôzne rozmiestnené po palubovke vozidla a na spätné zrkadlá. Kalibrácia červeného terču prebiehala v dvoch etapách. V prvej etape došlo ku kalibrácii v zafixovanej polohe hlavy vodiča, kedy vodič sleduje očami QR kódy. V druhej fáze kalibrácie vodič hýbe hlavou a sleduje pri tom určené miesta na palubovke, napríklad spätné zrkadlá. Následne sa z pozície očí a z pozície značiek dopočítava do videa, sprostredkovaného kamerou v eye-trackeri vodiča, červený terč, ktorý sa následne vykreslí, vid' Obrázok 7.2.

Neoddeliteľnou zložkou merania boli senzory v oboch pedáloch vozidla. Jeden snímал stlačenie rýchlостného pedálu a druhý senzor snímал stlačenie brzdového pedálu, pričom nasnímané informácie o polohe pedálov boli vynesené do grafu ako signál, vid' Obrázok 7.3.

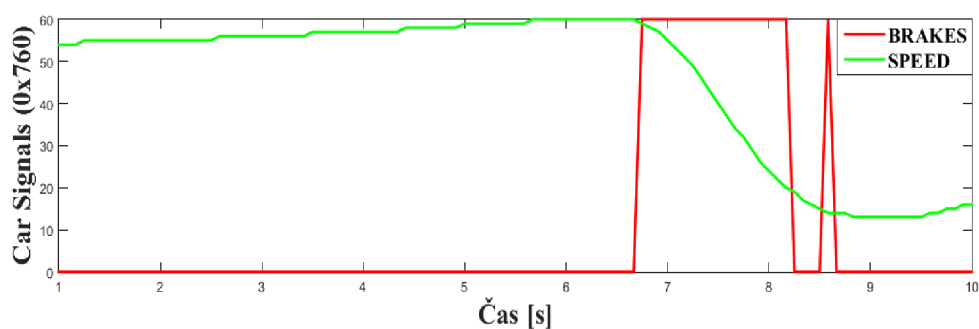
Na synchronizáciu dát bolo použité zariadenie D-Lab, ktoré bolo implementované v automobile a ktorý zabezpečuje jednotný štart všetkých zaznamenaných signálov, údajov z GPS a videozáznamu.



*Obrázok 7. 1 : Nasnímané polohy očí v jednom okamihu pomocou eyetrackera implementovaného v okuliaroch vodiča, naľavo ľavé oko, napravo pravé oko.*



Obrázok 7. 2 : Pohľad na interiér vozidla z implementovanej kamery v okuliaroch vodiča. Značky s "QR kódmi". Červený terč.



Obrázok 7. 3 : Záznam signálov z rýchlostného pedálu a brzdového pedálu, trvanie 10 sekúnd.

## 7.2 Vizualizácia dát

V rámci Ústavu biomedicínskeho inžinýrství bol vytvorený program v Pythone na vizualizáciu získaných dát z jazdných meraní. Odkaz na program: <https://github.com/branslavhesko/D-LAB-visualizer>. V programe je možné sledovanie jednej jazdy a všetkých akvizovaných dát nameraných práve počas tejto jazdy. Pre prehľadávanie dát je predovšetkým nutná implementácia relevantných dát od daného subjektu, ktorými sú biosignály, signály z brzdového a rýchlostného pedálu a údaje z GPS. V otvorenom programe je výber veľkosti časovej osi pre biosignály a signály z pedálov a výber veľkosti posunu okna. Ďalej nastavenie synchronizácie biosignálov, údajov zo stlačenia pedálov a samotného videa. Súčasťou programu je aj statický obraz aktuálneho časového úseku z kamery upevnenej na okuliaroch vodiča a mapa



s označovanou aktuálnou polohou vodiča na vyznačenej trase, ktorú vodič počas jazdy absolvoval. V spodnej časti programu sú znázornené dva grafy, jeden vykresľuje signály z brzdového a rýchlostného signálu a druhý biosignály, ktoré sa volia v checkboxoch v hornej časti programu. Program umožňuje multisignálové zobrazenie. Na posúvanie okna slúži tlačidlo umiestnené pod výberom veľkosti časovej osi alebo manuálne tlačidlo na posun o ľubovoľné okno umiestnený v dolnej časti. Program zabezpečuje ukladanie aktuálnych zobrazených dát do súborov s „.csv“ príponou. Na obrázku 7.4 je uvedená ukážka so spusteným programom.



*Obrázok 7. 4 : Príklad spusteného programu pre konkrétne dáta so zobrazeným statickým obrazom z kamery, aktuálnou polohou na mape, signálmi z pedálov a biosignálov so synchronizáciou.*

## 8. VÝBER RELEVANTNÝCH DÁT

Princíp výberu relevantných dát spočíval v prezeraní hodinových videozáznamov jazd a výberu úsekov, ktoré podľa subjektívneho názoru sledovateľa vykazovali známky stresu, kľudu, únavy a bez únavové úseky. Biosignály boli merané pomocou zariadenia BiosignalsPlux, opísaného v [24]. Vždy boli vybrané úseky dlhé 10 sekúnd. Každý zaznamenaný úsek jazdy bol extrahovaný z hodinového súboru signálov pomocou programu uvedeného v podkapitole 7.2 do súboru s koncovkou „.csv“ a spracovaný v programovacom prostredí Matlab, viď kapitola 9.

### 8.1 Dáta vykazujúce stres

V priebehu merania boli vodiči vystavení rôznym vonkajším podnetom, ktoré mohli vyvolať stres. Medzi faktory, ktoré spôsobujú stres, patrí aj celkový psychický a fyzický stav daného vodiča. Niektoré podnety boli vyvolané spontánne a niektoré boli plánované.

Medzi radu podnetov, ktoré boli vyvolané u subjektov, patrí napríklad náhle vykročenie chodca na priechod pre chodcov, vykročenie chodca na vozovku spoza nepriehľadnej rady áut, prudké brzdenie vozidla idúcom pred vodičom a jeho náhle vybočenie z cesty a predbiehajúce vozidlo v nepriehľadnej pravotočivej zákrute. Medzi ďalšie stresujúce motívy je možné zaradiť situáciu, kedy vodič obchádza radu áut, alebo cyklistov a oproti nemu sa blíži ďalšie vozidlo, alebo náhle zúženie vozovky. Stres môžu vyvolať aj zvieratá, v našom prípade konkrétne mačka a zajac, ktoré prechádzajú cez cestu.

Pri výbere dát zohrávalo dôležitú úlohu sledovanie záberov z kamery a výber relevantných úsekov z rady biosignálov, u ktorých bol predpoklad, že vykazujú zmeny z dôvodu vyvolaného stresu, so zameraním na vonkajší podnet. Signály boli vybrané na základe pozorovateľnej zmeny, respektíve výchylky od normálneho stavu. Podobne boli vybrané aj tie úseky jazdy, u ktorých sa predpokladalo, že v tom momente vodič nebol ovplyvnený žiadnymi vonkajšími podnetmi, ktorými sú chodci, predbiehajúce vozidlo, a podobne, teda úseky, ktoré zo subjektívneho hľadiska nevykazovali žiadne odchýlky. Tieto úseky biosignálov boli následne extrahované a spracované v programovacom prostredí Matlab. Niektoré signály boli pre spracovanie nevyhovujúce a preto sa v práci nevyskytujú. Príkladom je záznam EDA, ktorej signál bol často narušený šumom, alebo bolo zápästie vystavené slnečnému žiareniu a tak dochádzalo k nadmernému poteniu, čo nebolo príznakom zmeny stavu vodiča, a preto sa tento záznam považuje za irelevantný.

### 8.2 Dáta vykazujúce únavu

Pri výbere dát, ktoré vykazujú známky ospalosti a únavy zohráva dôležitú úlohu čas, v ktorom prebiehala jazda a doba trvania tejto jazdy. Denná jazda každého subjektu trvala

niečo vyše hodiny a nočná jazda trvala o niečo kratšie, väčšinou necelú hodinu. Nočná jazda prebiehala za zníženého výskytu áut na ceste a vypnutých semaforov, takže vodiči prešli rovnakú trasu za kratší čas oproti dennej jazde.

Pri dennej jazdy nebol predpoklad, že bude vykazovať známky únavy, ale vzhľadom k vysoko frekventovanej doprave v meste a obmedzeniach a zúženiach na diaľnici mohlo dôjsť na konci jazdy k vyčerpanosti vodiča a to mohlo byť podnetom pre únavu. Počas nočnej jazdy môže dôjsť k rýchlejšiemu vyčerpaniu, únave a to hlavne kvôli cirkadiánnemu rytmu, kedy je človek nastavený na spánok a podobne napríklad kvôli monotónnosti jazdy na diaľnici, kedy je výskyt áut obmedzenejší. Veľký vplyv na únavu má osvetlenie oproti idúcich áut, ktoré oslepujú vodiča.

## 9. SPRACOVANIE BIOSIGNÁLOV Z DATABÁZY

### 9.1 Načítanie dát

Okrem toho, že záznamy z jednotlivých jász boli rozdelené na 4 kategórie obsahujúce úseky jász stresové, kľudové, únavové, a bez únavové, tak všetky záznamy boli pridelené aj do komplexných súborov, kde 4 kategórie obsahovali záznamy príslušnej kategórie od jednotlivých jász. Dáta z úsekov jász boli uložené v súboroch „.csv“ a všetky tabuľky so vzorkami signálov obsahujúce biosignály boli načítané v prostredí Matlab, kde vzorky nameraných signálov boli uložené v stĺpcoch a počet riadkov odpovedal počtu vzoriek. Následne boli biosignály rozdelené do nových tabuliek podľa prislúchajúceho biosignálu, kde vzorky signálov boli uložené v riadkoch a počet riadkov odpovedal počtu extrahovaných úsekov pre danú kategóriu.

### 9.2 Predspracovanie dát

Všetky signály boli vzorkované so vzorkovacou frekvenciou  $f_{vz} = 1000$  Hz, ktorá bola takto nastavená, aby nedošlo k aliasingu, čiže k prekrytiu spektier a straty dôležitých informácií.

Prvá fáza predspracovania bola pre všetky biosignály rovnaká. Jednalo sa o úpravu dát podľa zadanej transformačnej funkcie od výrobcu zariadenia BiosignalsPlux. Príkladom je transformačná funkcia pre úpravu signálu EKG do požadovaného rozmedzia hodnôt (-1.47mV, +1.47mV) uvedeného v rovnici 9.1 a prevzatého z [47]. ADC je vzorkovaná hodnota signálu, VCC je prevádzkové napätie,  $G_{ECG}$  je zisk snímača,  $n$  je počet bitov kanála a ECG(mV) značí hodnotu signálu v milivoltch.

$$ECG(mV) = \frac{\left(\frac{ADC}{2^n} - \frac{1}{2}\right) \cdot VCC \cdot 1000}{G_{ECG}} \quad (9.1)$$

Každý biosignál sa upravuje podľa vlastnej transformačnej funkcie. Následne boli zo signálov odfiltrované nežiadúce narušenia ako drift, brum a obmedzenie vplyvu myopotenciálov opísané v podkapitolách 2.1 pre EKG, 2.2 pre EMG a 2.3 pre EDA.

Pre získanie príznakov z EKG je dôležité zvýraznenie QRS komplexu a detekcia R vln, ktorá bola opísaná tiež v podkapitole 2.1. Pre EMG signál je dôležité zvýrazniť frekvenčné pásmo, ktoré obsahuje najrelevantnejšie informácie. Postup uvedený v podkapitole 2.2. Signál EDA upravený pomocou transformačnej funkcie sa už ďalej neupravoval.

### **9.3 Výstup predspracovania**

Po tom ako boli jednotlivé biosignály uložené v tabuľkách, načítané v Matlabe a bola na nich urobená úprava, vid' podkapitola 9.2, boli výsledné upravené signály uložené do novej tabuľky. Výstupom predspracovania teda boli nové „.csv“ súbory s predspracovanými biosignálmi, kde riadky obsahovali opäť vzorky signálu a počet riadkov odpovedal počtu extrahovaných úsekov pre danú kategóriu. Počet biosignálov, ktoré vykazovali normálny stav vodiča bolo 98, stresových stavov bolo 81, tých čo vykazovali stav bez únavy bolo 56 a únavových bolo 72.

## 10. EXTRAKCIA RELEVANTNÝCH PRÍZNAKOV

Extrakcia príznakov je dôležitý proces pre získanie modelu, ktorý produkuje spoľahlivé a kvalitné výsledky. Použitie nevhodných príznakov má za následok nepresný model produkujúci zavádzajúce výsledky. Preto je extrakcia príznakov dôležitým krokom v metódach strojového učenia.

### 10.1 EKG

Biosignál EKG je pravdepodobne najviac preštudovaným biologickým signálom a preto ponúka širokú škálu príznakov potrebných pre zostrojenie klasifikačného modelu. Pre získanie väčšiny uvedených príznakov bolo potrebné uskutočniť detekciu R vln. Postup bol uvedený v podkapitole 2.1. Medzi príznaky vypočítané z detekovaných R vln sa zaraďujú minNN, maxNN, MNN, SDNN, RMSSD, NN20, NN50, pNN20, pNN50, ktoré budú vysvetlené v nasledujúcej tabuľke a boli počítané podľa [48], ďalej Triangular index HRV a TINN vypočítané podľa [25], a HR podľa [49]. Niektoré príznaky ako je LF, HF a pomer LF/HF boli najprv vypočítané a následne vypustené z práce, pretože tieto príznaky sa používajú na záznamy dlhé 2 minúty a v práci boli použité záznamy dlhé 10 sekúnd.

Tabuľka 10. 1 : Extrahované príznaky zo signálu EKG.

<b>Príznak</b>	<b>Popis</b>
<b>minNN</b>	<i>Minimálna hodnota rozdielov NN intervalov</i>
<b>maxNN</b>	<i>Maximálna hodnota rozdielov NN intervalov</i>
<b>MNN</b>	<i>Priemerná hodnota NN intervalov</i>
<b>SDNN</b>	<i>Smerodajná odchýlka priemerných NN intervalov</i>
<b>RMSSD</b>	<i>Odmocnina priemerov kvadrátov diferencií susedných NN intervalov</i>
<b>HR</b>	<i>Tep srdca</i>
<b>NN20</b>	<i>Počet párov susedných NN intervalov líšiacich sa viac než o 20 ms</i>
<b>NN50</b>	<i>Počet párov susedných NN intervalov líšiacich sa viac než o 50 ms</i>
<b>pNN20</b>	<i>Počet NN20 delený celkovým počtom NN intervalov vynásobený 100</i>
<b>pNN50</b>	<i>Počet NN50 delený celkovým počtom NN intervalov vynásobený 100</i>
<b>Triangular index</b>	<i>Integrál hustoty histogramu NN intervalu podelený jeho výškou</i>
<b>TINN</b>	<i>Šírka základne najlepšej trojuholníkovej aproximácie histogramu</i>

V tabuľke 10.2 sú znázornené príklady extrahovaných príznakov z nočnej jazdy jedného zo subjektov pre 4 kategórie (TI = Triangular index).

Tabuľka 10. 2 : Príklady extrahovaných príznakov pre EKG.

	minNN	maxNN	MNN	SDNN	RMSSD		pNN50	TI	TINN
<b>Kľud</b>	575	600	587,62	7,78	5,38	...	0	3,2	0,03
<b>Stres</b>	494	545	514,06	13,11	9,45		0	4,25	0,05
<b>Bez únavy</b>	519	546	535,61	9,09	5,25		0	2,57	0,03
<b>Únava</b>	838	1030	930,9	52,53	60,54		50	5	0,20

## 10.2EMG

Príznaky extrahované z biologického signálu EMG zohrávajú dôležitú úlohu pre detekciu stresu a únavy. Niektoré signály boli extrahované v časovej oblasti a medzi nich patrí Energy, RMS, MAV, Variance, a niektoré vo frekvenčnej oblasti a to MNF, MDF, ktoré boli vypočítané podľa [50] a PowerSignal, vypočítaný podľa postupu uvedeného v [51]. V tabuľke 10.3 sú uvedené popisy jednotlivých príznakov a v tabuľke 10.4 sú uvedené príklady extrahovaných príznakov z nočnej jazdy jedného subjektu.

Tabuľka 10. 3 : Extrahované príznaky zo signálu EMG.

<b>Príznak</b>	<b>Popis</b>
<b>Energy</b>	<i>Energia signálu</i>
<b>RMS</b>	<i>Stredný priemer štvorca</i>
<b>MAV</b>	<i>Stredná absolútna hodnota signálu</i>
<b>Variance</b>	<i>Rozptyl signálu</i>
<b>MNF</b>	<i>Priemerná frekvencia</i>
<b>MDF</b>	<i>Stredná frekvencia</i>
<b>PowerSignal</b>	<i>Výkon signálu</i>

Tabuľka 10. 4 : Príklady extrahovaných príznakov z EMG.

	Energy	RMS	MAV	Variance	MNF	MDF	PowerSignal
<b>Kľud</b>	0,4701	0,01	0,01	4,7e-05	153,68	0,70	3,41e-05
<b>Stres</b>	78,88	0,09	0,07	0,01	118,43	0,63	0,01
<b>Bez únavy</b>	4,32	0,02	0,1	4,32e-04	154,82	0,86	6,66e-04
<b>Únava</b>	10,86	0,3	0,02	0	141,62	0,73	6,06e-04

## 10.3 EDA

Signál EDA neposkytoval vo veľa prípadoch požadované hodnoty. Signál bol veľa krát postihnutý zákmitmi v podobe QRS komplexov, alebo boli elektródy zle umiestnené a došlo k zašumeniu, alebo vykazuje známky saturácie. Preto tieto signály neboli použité v opisnej štatistike ani v klasifikačných modeloch. Najlepšie signál EDA vystihuje jeho stredná hodnota, ktorá bola zvolená za hlavný príznak [28]. Ďalšie jednoduchšie príznaky, ktoré boli odvodené sú Minimum, Maximum, Smerodajná odchýlka a Rozptyl, ktoré tvoria základný štatistický popis signálu. Popis príznakov je uvedený v nasledujúcej tabuľke a príklady vypočítaných príznakov z nočnej jazdy jedného subjektu sú uvedené v tabuľke 10.6.

Tabuľka 10. 5 : Extrahované príznaky zo signálu EDA.

<b>Príznak</b>	<b>Popis</b>
<b>Stredná hodnota</b>	<i>Priemerná hodnota</i>
<b>Minimum</b>	<i>Najmenšia hodnota</i>
<b>Maximum</b>	<i>Najväčšia hodnota</i>
<b>Smerodajná odchýlka</b>	<i>Kladná druhá odmocnina z rozptylu</i>
<b>Rozptyl</b>	<i>Stredná kvadratická odchýlka od výberového priemeru</i>

Tabuľka 10. 6 : Príklady extrahovaných príznakov z EDA.

	<b>Stredná hodnota</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Smerodajná odchýlka</b>	<b>Rozptyl</b>
<b>Kľud</b>	0,25	0,03	0,67	0,36	0,13
<b>Stres</b>	0,71	0,68	0,75	0,04	0
<b>Bez únavy</b>	0,27	0,04	0,50	0,32	0,10
<b>Únava</b>	0,52	0,50	0,54	0,04	0

## 10.4 Výstup extrakcie príznakov

Príznaky boli extrahované z každého biosignálu zvlášť a boli uložené do nových tabuliek, kde v riadku boli uložené príznaky extrahované z daného úseku jazdy za sebou ako sú napríklad uvedené v tabuľkách 10.2, 10.4 a 10.6, a počet riadkov odpovedal počtu extrahovaných úsekov pre danú kategóriu. Pre vytvorenie klasifikačného modelu boli príznaky z EKG a EMG prislúchajúceho úseku jazdy a kategórie spojené za sebou do novej tabuľky. Výstupná tabuľka s extrahovanými príznakmi pre stres obsahovala 179 meraní (pozorovaní) a pre únavu pozostávala zo 128 meraní.



# 11. ELIMINÁCIA DÁT A SELEKCIA PRÍZNAKOV

Pri veľkom súbore dát je možné, že niektoré prvky budú obsahovať údaje, ktoré nie sú vhodné pre ďalšie spracovanie. Preto je potrebné urobiť elimináciu týchto dát, aby výsledok práce dosahoval čo najlepšie výsledky. Ďalším krokom je selekcia príznakov, pretože nie všetky príznaky prinášajú do vznikajúceho modelu vysokú váhu pre správnu klasifikáciu.

## 11.1 Eliminácia nevhodných dát

Po výpočte príznakov pre uvedené biologické signály boli v tabuľke s príznakmi detekované prvky s „NaN“, tzv. nečíslo a „Inf“, teda nekonečno, ktoré by mohli spôsobovať problémy a preto boli nahradené 0.

Následne bolo zistené, že pri niektorých záznamoch EKG boli vlny R nesprávne detekované, a to spôsobilo vznik veľmi odľahlých hodnôt ( $\text{minNN} = 2$ ;  $\text{maxNN} = 9100$ ). Tieto hodnoty nekorešponujú s požadovanými a preto bolo celé dané meranie odstránené.

Pri ďalšej analýze sa zisťovalo, či jednotlivé merania, alebo jednotlivé príznaky cez všetky merania, neobsahujú počet núl prevyšujúci 30% z celkového počtu. Všetky merania boli zachované, ale príznaky NN50 a pNN50 cez všetky merania obsahovali počet núl prevyšujúcich 30% a preto boli z finálnej tabuľky príznakov odstránené.

Výstupom z eliminácie dát a vstupom pre selekciu príznakov bola tabuľka veľkosti 168x17 pre stres a 117x17 pre únavu, kde prvé číslo definovalo počet meraní a druhé číslo definovalo počet príznakov.

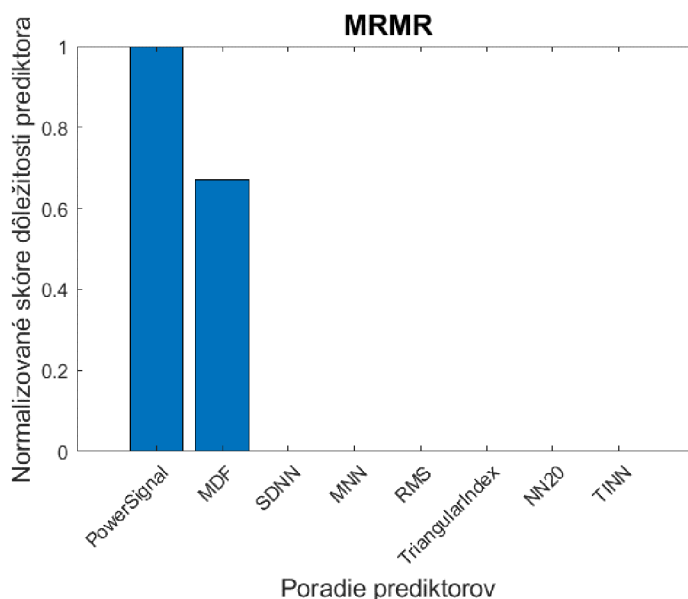
## 11.2 Selekcia príznakov pre klasifikáciu stresu

Selekcia príznakov bola uskutočnená 4 technikami a vybraná bola tá, ktorá po testovaní natrénovaného modelu dala najuspokojivejší výsledok.

Prvou základnou metódou, alebo skôr metrikou na meranie závislosti bola korelácia medzi príznakmi a vyhľadanie tých, ktoré medzi sebou korelovali o viac než 0,95. Korelácia bola vysvetlená v podkapitole 3.3. Ukázalo sa, že ak bol odstránený jeden z dvojice príznakov, ktoré spolu korelovali, tak model začal vykazovať horšie výsledky oproti ostatným selekčným metódam, preto sa táto metóda zamietla.

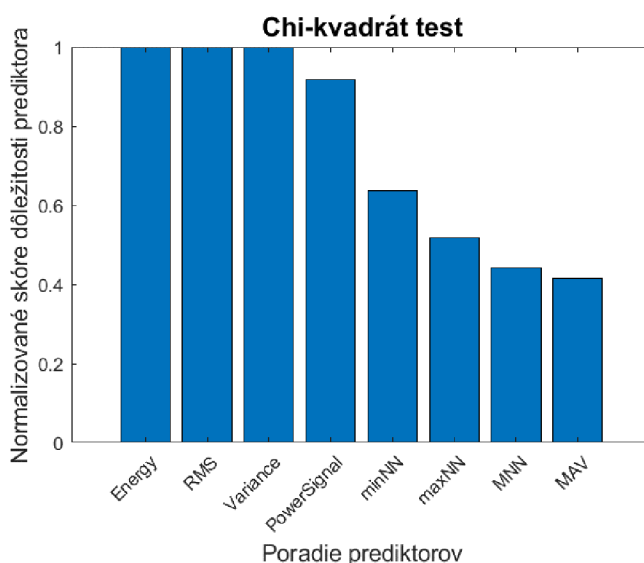
Ďalšou metódou, ktorá bola použitá bol algoritmus MRMR, ktorý bol vysvetlený v podkapitole 3.4. Po uskutočnení tréovania a testovania modelu, ktorého príznaky boli selektované pomocou tohto algoritmu, bola úspešnosť modelu omnoho vyššia ako pri selekcii príznakov podľa vzoru korelácie. Stĺpcový graf 11.1, ktorý znázorňuje poradie prediktorov podľa skóre dôležitosti príznakov určuje, že len príznaky PowerSignal a MDF sú dôležité a dosahujú vysoké hodnoty. Ostatné príznaky daný algoritmus

neoznačil za dôležité a pridelil im veľmi nízke hodnoty. V grafe je znázornených 8 najdôležitejších príznakov, ktoré boli použité aj v klasifikačných modeloch.



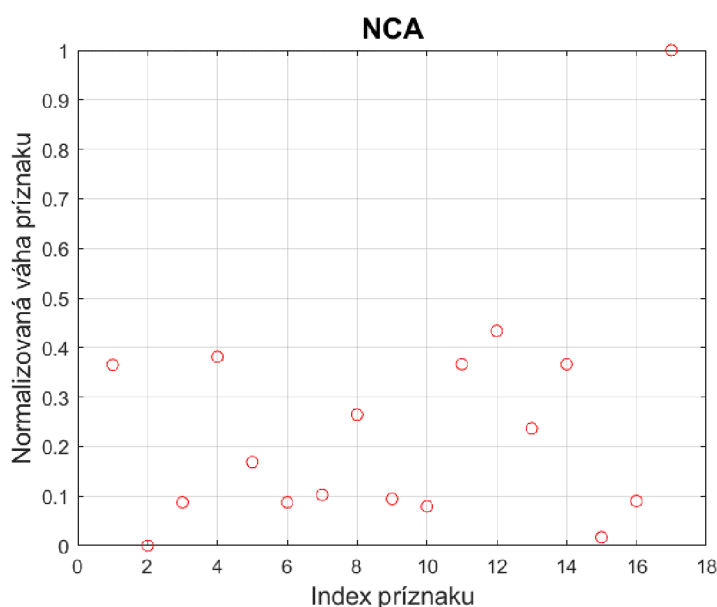
Graf 11. 1 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa algoritmu MRMR pre stres.

Na meranie závislosti príznakov sa môžu použiť aj štatistické testy a jedným z nich je chi-kvadrát test, popísaný v podkapitole 3.3, ktorý bol použitý na selekciu príznakov a vďaka nemu dosahoval výsledný klasifikačný model veľmi priaznivé výsledky. Nasledujúci stĺpcový graf 11.2 vyznačuje zostupné poradie 8 najlepších príznakov podľa určeného skóre dôležitosti, ktoré boli použité aj v klasifikačných modeloch.



Graf 11. 2 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa chi-kvadrát testu pre stres.

Algoritmus NCA, ktorý analyzuje susedské komponenty pridelil príznaku PowerSignal najvyššiu váhu a dvojnásobnú váhu oproti druhému najlepšiemu príznaku, vid' graf 11.3. Najmenšiu významnosť priradil príznaku maxNN a druhú najnižšiu MNF. Naproti tomu chi-kvadrát test určil príznak maxNN ako dôležitý. V tabuľke 11.1 je vybraných prvých 8 príznakov, ktorým NCA udelil najvyššiu váhu. Pri porovnávaní vybraných príznakov pomocou NCA s vybranými príznakmi pomocou chi-kvadrát testu došlo u 5 príznakov k zhode. Jedná sa o Energy, RMS, Variance, PowerSignal a minNN. Z toho je možné usúdiť, že dané príznaky sú pre návrh klasifikačného modelu dôležité. Tento algoritmus sa ako jediný mohol použiť aj pre klasifikačnú metódu logistická regresia, pričom vybral rovnaké príznaky aj po úprave parametrov.



Graf 11. 3 - Normalizované váhy príznakov udelených algoritmom NCA pre stres.

Tabuľka 11. 1 : Prvých 8 príznakov s najvyššími váhami určenými algoritmom NCA pre stres.

<b>NCA</b>	minNN	SDNN	pNN20	Energy	RMS	MAV	Variance	PowerSignal
------------	-------	------	-------	--------	-----	-----	----------	-------------

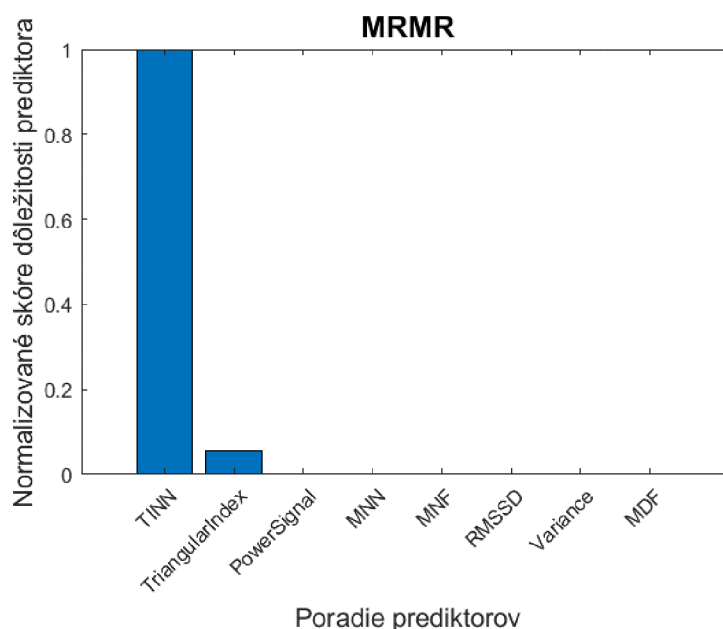
### 11.3 Výber najlepšej selekčnej metódy pre klasifikáciu stresu

Všetky selekčné metódy boli postupne aplikované na extrahované príznaky a klasifikačné modely boli natréňované na trénovacej množine a overené na testovacej, pri ktorej sa počítala presnosť a AUC hodnota. Následne boli jednotlivé selekčné metódy porovnané na základe výsledkov presnosti a AUC hodnoty a vybrané príznaky pomocou algoritmu NCA sa osvedčili ako najvhodnejšia kombinácia na dosiahnutie najlepších výsledkov

klasifikácie. Preto bola táto selekčná metóda zvolená pre všetky klasifikačné modely. Vybraná bola kombinácia príznakov uvedená v tabuľke 11.1, ktoré boli určené na základe prahu definujúceho výber tých príznakov, ktorých váha je väčšia ako 0,2.

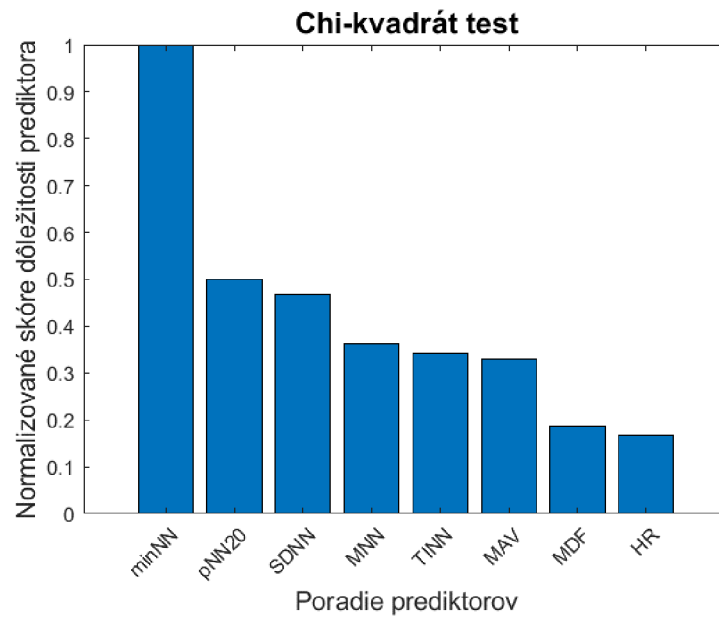
## 11.4 Selekcia príznakov pre klasifikáciu únavy

Ako bolo uvedené v podkapitole 11.2, korelácia sa neosvedčila ani pre klasifikáciu únavy ako dobrá metóda na selekciu príznakov. Algoritmus MRMR sa použil aj na výber vhodných príznakov pre klasifikáciu únavy. Výsledok je zobrazený v grafe 11.4. Opäť algoritmus MRMR vybral len dva príznaky, ktoré sú podľa neho najdôležitejšie, pričom druhý príznak dosahuje omnoho menšie skóre, než prvý prediktor. Ďalšie príznaky dosahovali skóre v  $e-14$  až 0.



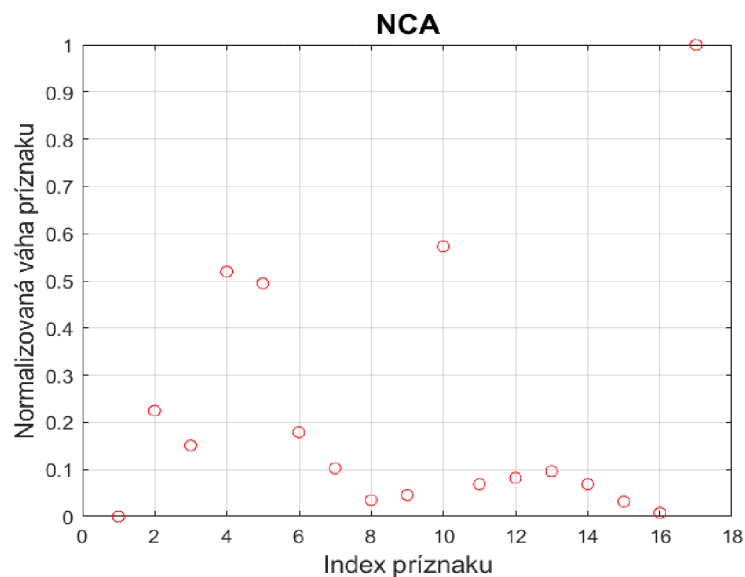
Graf 11. 4 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa algoritmu MRMR pre únavu.

Chi-kvadrát test sa použil na výber vhodných príznakov aj v prípade dát pre klasifikáciu únavy a jeho výsledok je znázornený v grafe 11.5. V tomto prípade pre klasifikáciu únavy boli vybrané iné príznaky ako pri selekcii príznakov pre stres. Na ose, kde je znázornené poradie prediktorov (príznakov) je vidieť, že najdôležitejší príznak je v tomto prípade minNN, ktorý bol pri strese zaradený až ako 4.tý dôležitý príznak.



Graf 11. 5 : Skóre dôležitosti prediktorov a ich zoradenie od najvyššieho skóre podľa chi-kvadrát testu pre únavu.

Poslednou metódou selekcie aplikovanou na sadu príznakov pre klasifikáciu únavy bol algoritmus NCA, ktorého výsledné normalizované váhy sú znázornené v nasledujúcom grafe. Príznak PowerSignal bol znova vybraný za príznak s najvyššou váhou a príznak minNN podobne za príznak s najnižšou váhou ako pri selekcii príznakov pre klasifikáciu stresu. Tabuľka 11.2 obsahuje prvých 8 vybraných príznakov. Podobne bol tento algoritmus použitý aj na logistickú regresiu a výsledkom bol výber rovnakých príznakov.



Graf 11. 6 : Normalizované váhy príznakov udelených algoritmom NCA pre únavu.

Tabuľka 11. 2 : Prvých 8 príznakov s najvyššími váhami určenými algoritmom NCA pre únavu.

NCA	maxNN	MNN	SDNN	RMSSD	HR	NN20	TINN	PowerSignal
-----	-------	-----	------	-------	----	------	------	-------------

### 11.5 Výber najlepšej selekčnej metódy pre klasifikáciu únavy

V procese rozhodovania o najlepšej selekčnej metóde boli na každý klasifikačný model použité vyššie uvedené metódy a zisťovalo sa, ktorá metóda má najvýznamnejší príspevok v zostrojení klasifikačného modelu, ktorý produkuje vysokú presnosť a veľkosť AUC. V tomto procese bola opäť vybraný algoritmus NCA, pretože pri jej použití dosahovali všetky klasifikačné modely najlepšie výsledky. Vybraná bola kombinácia príznakov uvedená v tabuľke 11.2, ktoré boli určené na základe prahu, ktorý definuje výber tých príznakov, ktorých váha je väčšia ako 0,1.

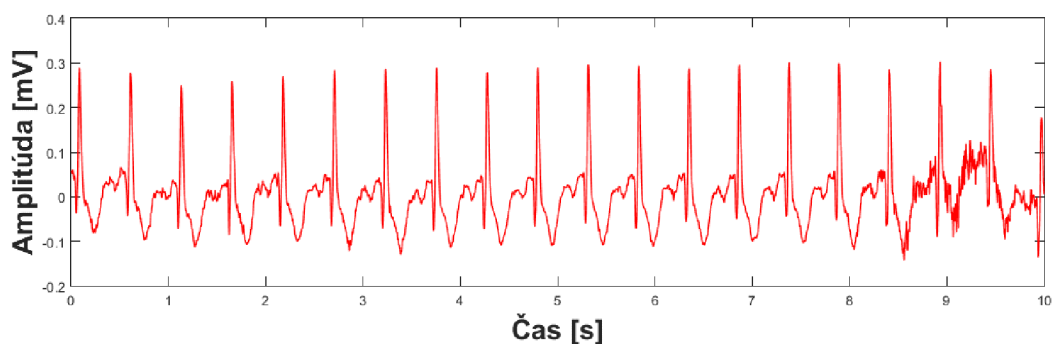
## 12. PRÍKLAD NAMERANÝCH BIOSIGNÁLOV

V nasledujúcich prípadoch sú zobrazené také príklady biologických signálov, ktoré najlepšie vystihujú signály potrebné pre túto prácu, ale aj také, ktoré boli zle namerané a sú pre ďalšiu prácu nepoužiteľné. Nasledujúce obrázky tejto kapitoly slúžia ako ukážka.

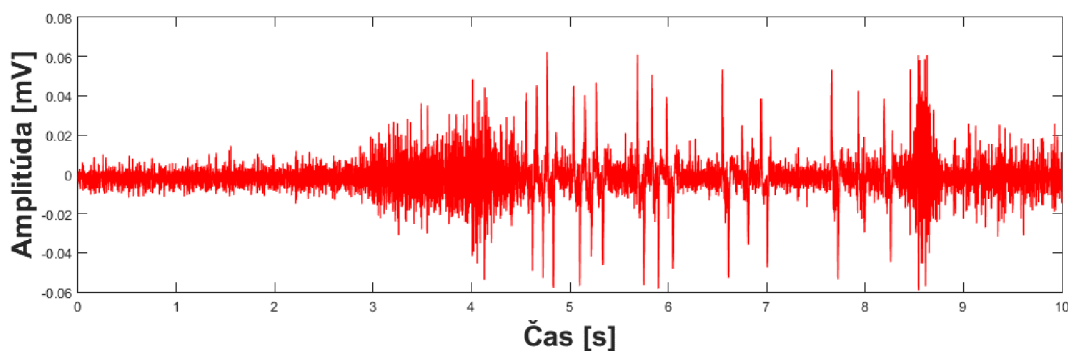
### 12.1 Úseky vykazujúce normálny stav a stresový stav

- Prvý prípad

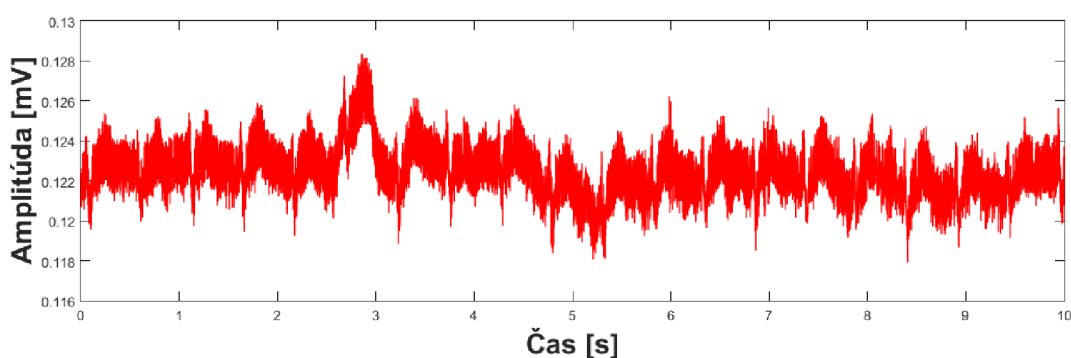
K stresu môže dôjsť z rôznych príčin. Vonkajší faktor, ktorý pravdepodobne vyvolal stres v tomto prípade je zúženie v prechode pod železničným mostom, kedy došlo k styku s protiidúcim autom, pričom vodič toto auto zbadal neskoro. Z obrázka 12.1, ktorý znázorňuje EKG je pozorovateľná zmena signálu s počiatkom približne medzi 8 a 9 sekundou, čo korešponduje presne s pozorovaním protiidúceho auta. Na obrázku 12.2, ktorý vykresľuje EMG, je vidieť, že došlo v tom istom čase k nárastu amplitúdy, čo značí kontrakciu svalu. Kontrakcia svalstva môže byť zapríčinená aj vybudením stresu z pokojného stavu vodiča. Približne v tretej sekunde došlo podobne ku kontrakcii svalstva, pravdepodobne z inej príčiny, pretože záznam videa nepreukázal dôvod k vyvolaniu stresu. V tomto prípade sa záznam EDA prekrýva so záznamom EKG, viď obrázok 12.3, kedy pravdepodobne došlo k umiestneniu elektród ďalej od seba a elektródy zachytili aj srdečný signál v podobe zákmitov, ktoré značia QRS komplex a sú totožné s QRS komplexmi EKG signálu, uvedeného na obrázku 12.1.



Obrázok 12. 1 : Biosignál EKG, trvanie 10s.



Obrázok 12. 2 : Biosignál EMG, trvanie 10s.

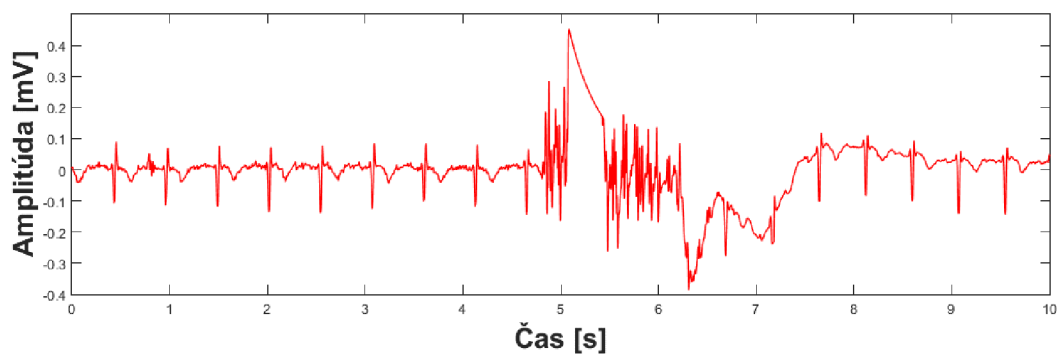


Obrázok 12. 3 : Biosignál EDA, trvanie 10s.

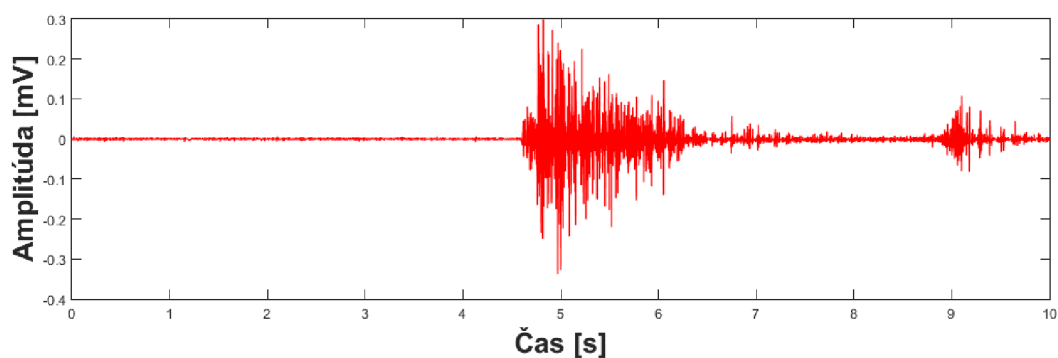
- **Druhý prípad**

Nasledujúce obrázky ponúkajú veľmi dobrý príklad toho, že so zmenou z pokojného stavu vodiča do stresového stavu vodiča dochádza k zmene v biosignáloch. V tomto prípade bola vonkajším faktorom lopta, ktorá sa skotúľala na cestu z poza auta. Cesta predtým bola pokojná a signály po dobu 5 minút pred incidentom nevykazovali žiadne známky stresu. Na uvedených obrázkoch, po sebe idúcich, je pozorovateľná zmena u EKG a EMG tesne pred začatím 5 sekundy a u EDA tesne po začatí 5 sekundy. V prípade EKG došlo k úplnému rozrušeniu signálu, u ktorého nie je viditeľný žiaden QRS komplex po dobu 2,5 sekundy. U EMG je očividný nárast amplitúdy signálu, čo znamená kontrakciu svalstva, a EDA vykazuje prudký prírastok vodivosti, ku ktorej dôjde v prípade zvýšenej produkcie potu, čo je jeden z príznakov stresu [21].

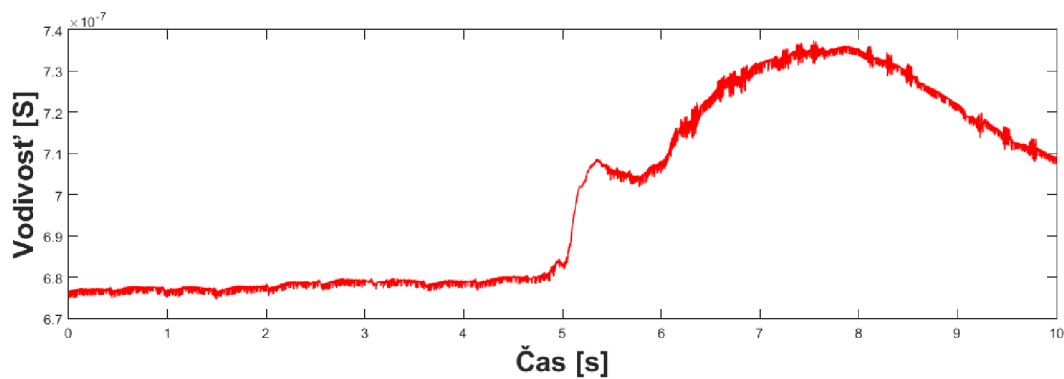




Obrázok 12. 4 : Biosignál EKG, 10s.



Obrázok 12. 5 : Biosignál EMG, 10s.

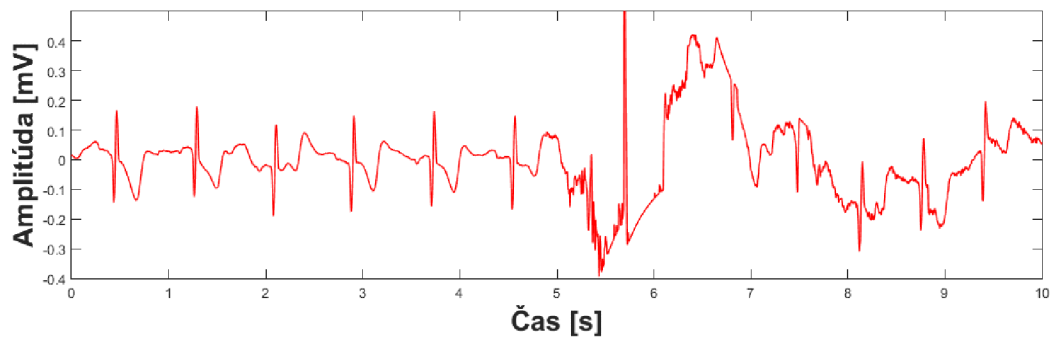


Obrázok 12. 6 : Biosignál EDA, 10s.

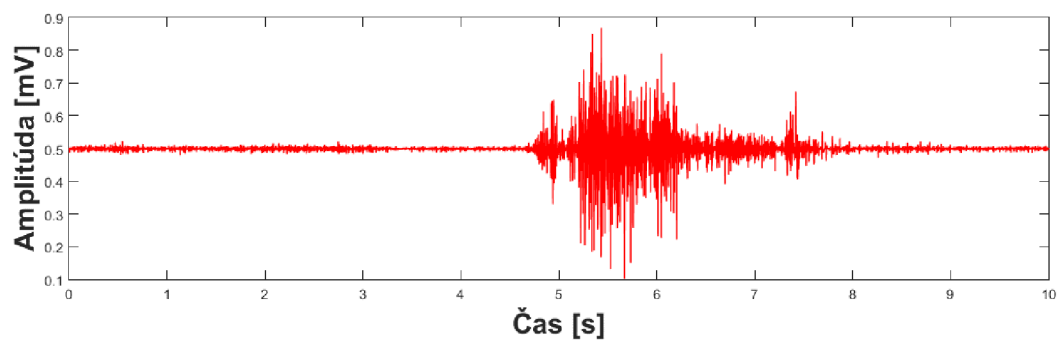
- **Tretí prípad**

Pri jazde v meste môže dôjsť k rôznym situáciám, ktoré vyvolajú u vodiča stres. Jednou z príčin sú chodci, ktorí vstupujú na vozovku bez toho, aby sa uistili, že ich žiadne auto neohrozuje. Signál EKG, zobrazený na obrázku 12.7, vykazuje zmenu tesne pred začatím piatej sekundy, kedy došlo k úplnej deštrukcii signálu, ktorý po dobu troch sekúnd

neobsahuje žiadne QRS komplexy. U signálu EMG nastala podobná situácia tesne pred piatou sekundou v podobe kontrakcie svalstva a amplitúda sa znížila až po troch sekundách.



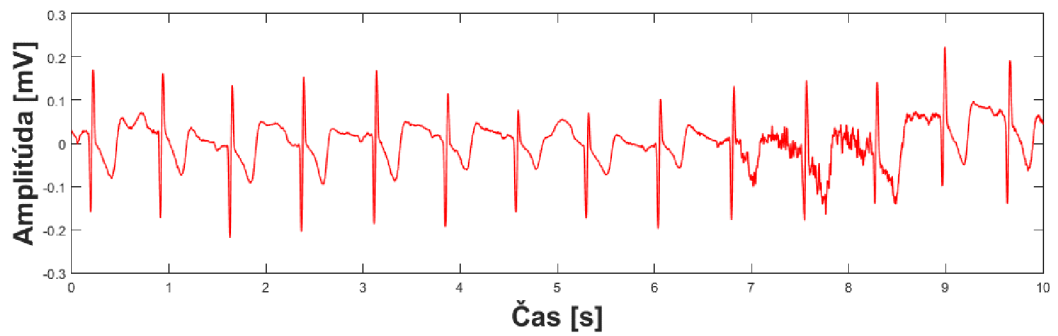
Obrázok 12. 7 : EKG biosignál, 10s.



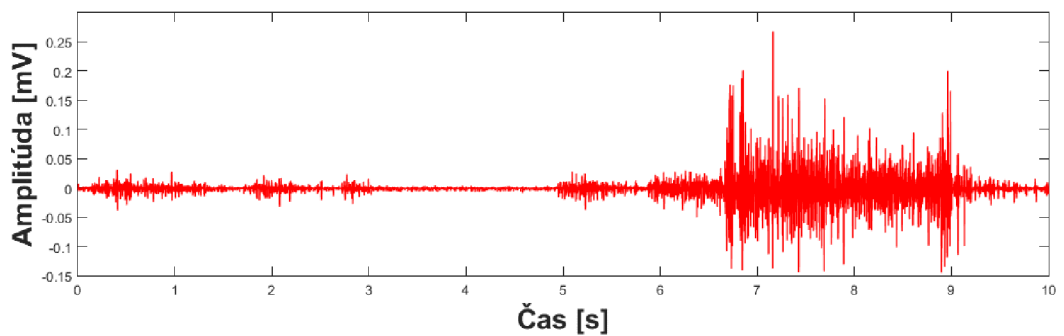
Obrázok 12. 8 : EMG biosignál, 10s.

- **Štvrtý prípad**

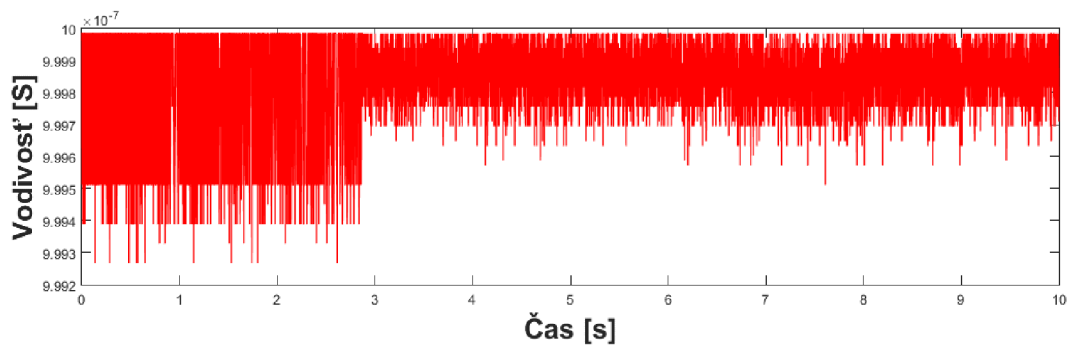
Posledný prípad uvedený v tejto kapitole vystihuje situáciu, kedy auto idúce pred vodičom náhle prudko zabrzdilo a zmenilo smer jazdy. Táto situácia nastáva nezriedka a vodič nie je schopný túto akciu predpovedať. Preto dochádza k pocitovej zmene u vodiča z kľudného do vypätého stavu. U biosignáloch EKG a EMG je rozoznateľná zmena pred počiatkom siedmej sekundy, v zmysle odvrátenia od normálneho stavu. Záznam EDA, obrázok 12.11, vykazuje známky saturácie, a preto sa tento signál považuje za neuspokojivý a nepostačujúci pre ďalšiu prácu a bol z merania odobraný.



Obrázok 12. 9 : EKG biosignál, 10s.



Obrázok 12. 10 : EMG biosignál, 10s.



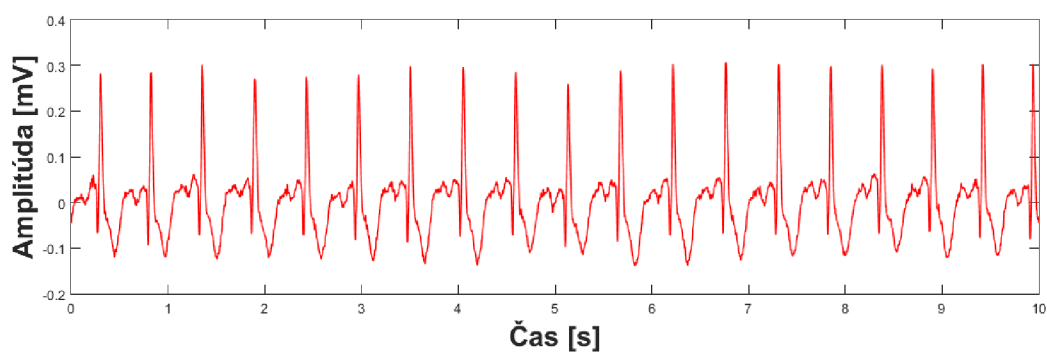
Obrázok 12. 11 : EDA biosignál - nevyhovujúci, 10s.

## 12.2 Úseky bez únavy a vykazujúce únavu

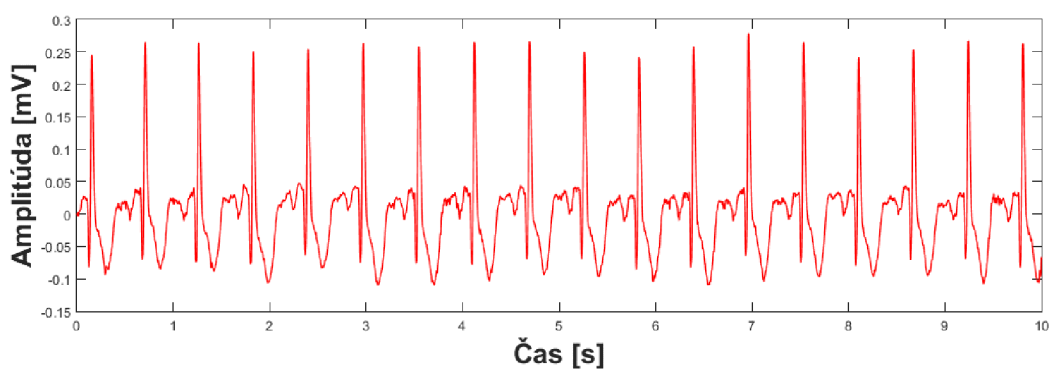
- **Denná jazda**

Ako prvé sú znázornené signály EKG. Už pri prvom pohľade je evidentné, že signály nevykazujú žiadne odchýlky, žiadne abnormality. Pred každým komplexom QRS sa nachádza P vlna a za každým komplexom nasleduje T vlna. Medzi obrázkom 12.12, kde

ja záznam EKG na začiatku jazdy a obrázkom 12.13, kde je záznam EKG pred ukončením jazdy nie sú veľké rozdiely. Únava zapríčiňuje mierne zníženie priemerného tepu srdca. Pri detailnejšom skúmaní bolo zistené, že skoro u všetkých subjektov bola znížená srdečná frekvencia v úseku pred ukončením jazdy. Tento poznatok bol zistený zrátaním R vln, ktorých počet po uplynutí jazdy poklesol. V tejto časti práce neprebehla ešte žiadna detailná analýza, zaznamenaný bol len subjektívny názor.

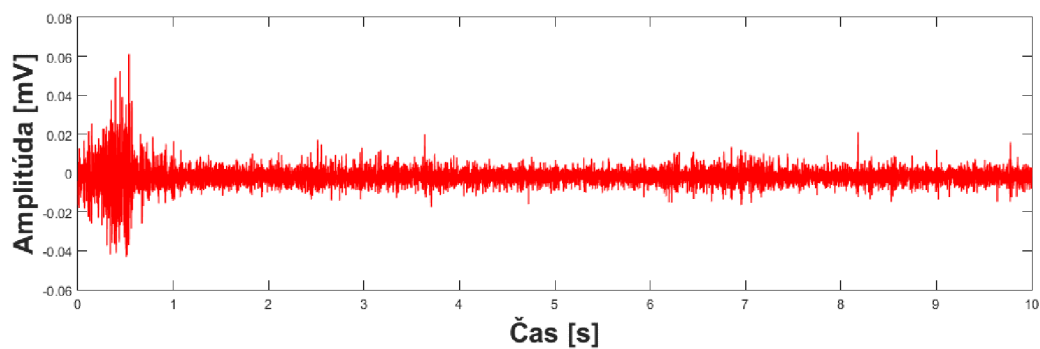


*Obrázok 12. 12 : EKG biosignál, začiatok dennej jazdy, 10s.*

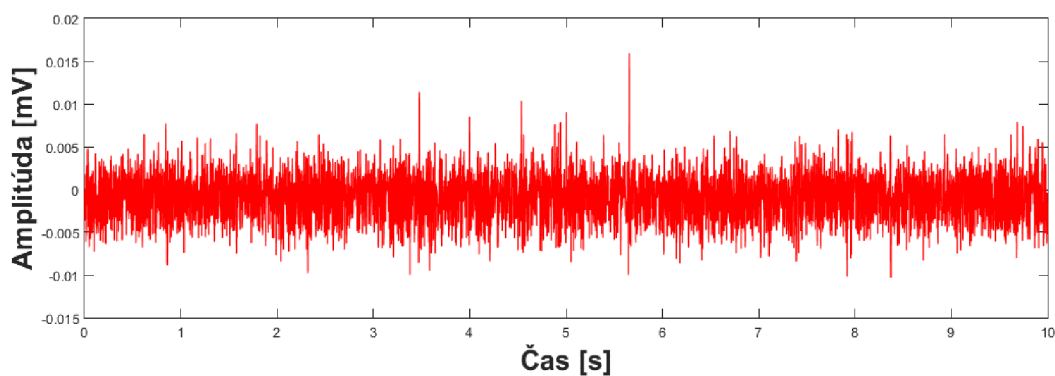


*Obrázok 12. 13 : EKG biosignál, koniec dennej jazdy, 10s.*

Nasledujúce dve obrázky zachytávajú priebeh biosignálu EMG na začiatku a na konci jazdy. V originálnej oblasti je ťažko posúdiť, či po uplynutí jazdy bola u vodiča navodená únava. Aby bola potvrdená, alebo vyvrátená únava u signáloch EMG, bude potrebné vykonať popisnú štatistiku, čo sa uskutoční v kapitole 13.

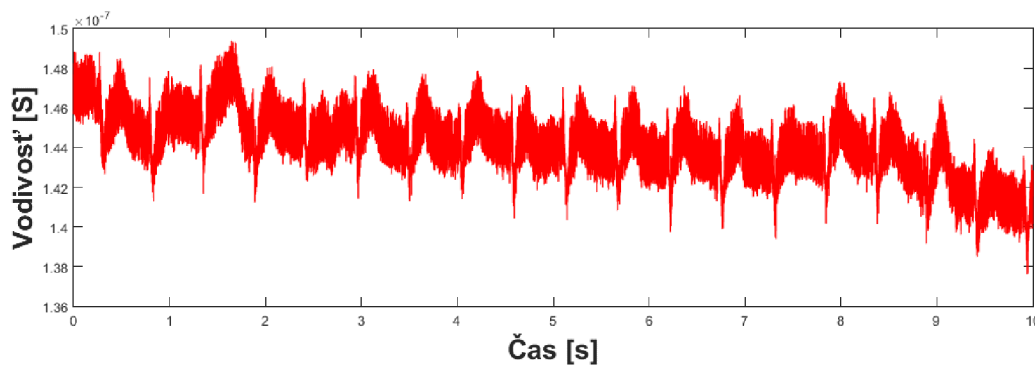


*Obrázok 12. 14 : EKG biosignál, koniec dennej jazdy, 10s.*

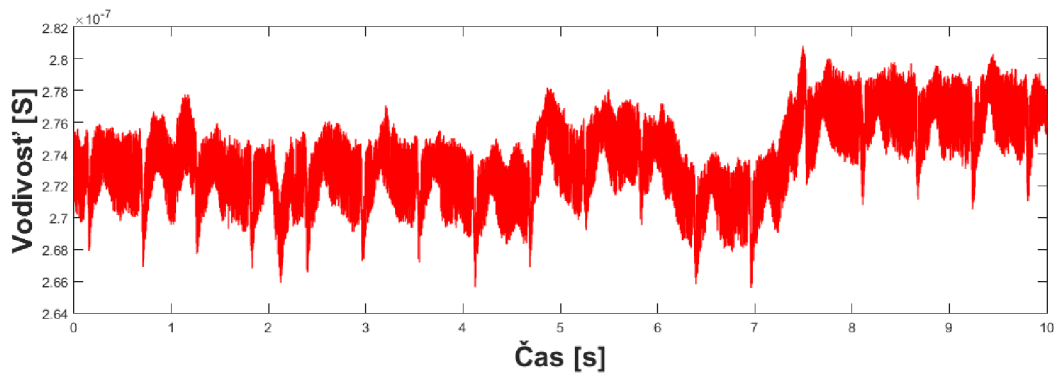


*Obrázok 12. 15 : EMG biosignál, koniec dennej jazdy, 10s*

Na obrázkoch 12.16 a 12.17 sú záznamy biosignálov EDA, v ktorých sú viditeľné zákmity QRS komplexov, ktoré korelujú so signálom EKG vykresleným na obrázkoch 12.12 a 12.13.



*Obrázok 12. 16 : EDA biosignál, začiatok dennej jazdy, 10s.*

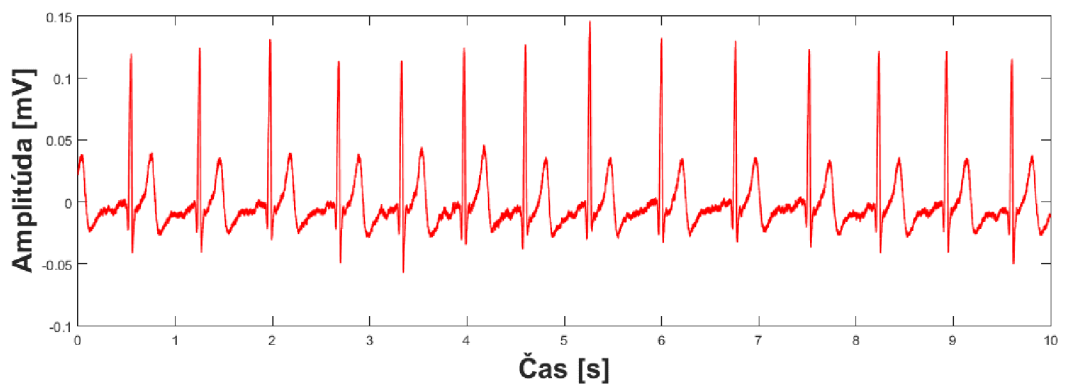


Obrázok 12. 17 : EDA biosignál, koniec dennej jazdy, 10s.

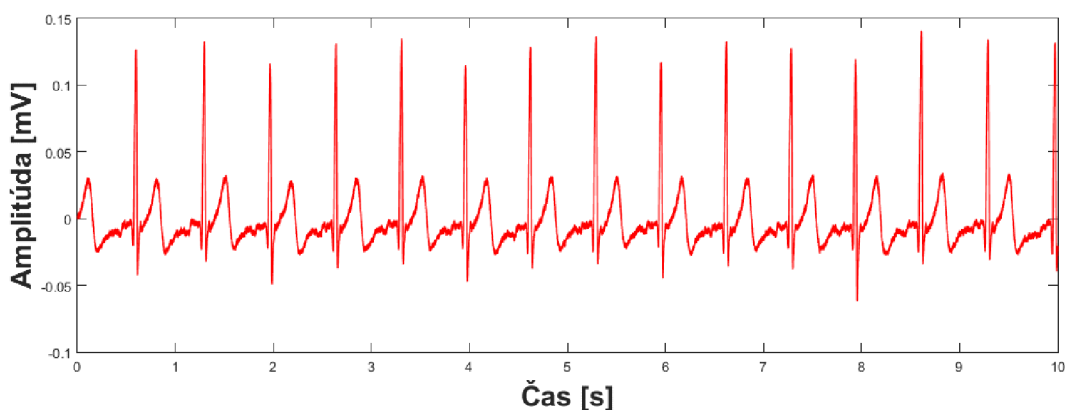
- **Nočná jazda**

Pred nočnou jazdou subjekty podstúpili záťažový test na počítači, ktorý pozostával zo súboru úloh na rôzne mentálne aktivity, ako pamäťové, matematické, a pod. Tento test bol vytvorený, aby unavil vodiča pred jazdou.

Medzi signálmi EKG v originálnej oblasti, na začiatku jazdy a na konci jazdy, neboli zistené veľké rozdiely. V porovnaní s výsledkami z dennej jazdy, kde bol zistený pokles srdečnej frekvencie u väčšiny subjektov pred ukončením jazdy oproti začiatku jazdy, bol pokles srdečnej frekvencie definovaný len u 50 % subjektov. Iné rozdiely možno preukáže detailnejšia analýza v nasledujúcej kapitole tejto práce.

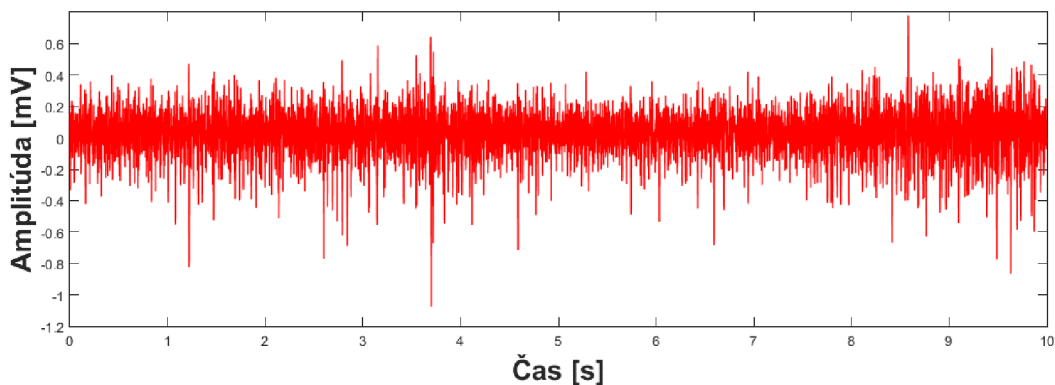


Obrázok 12. 18 : EKG biosignál, začiatok nočnej jazdy, 10s.

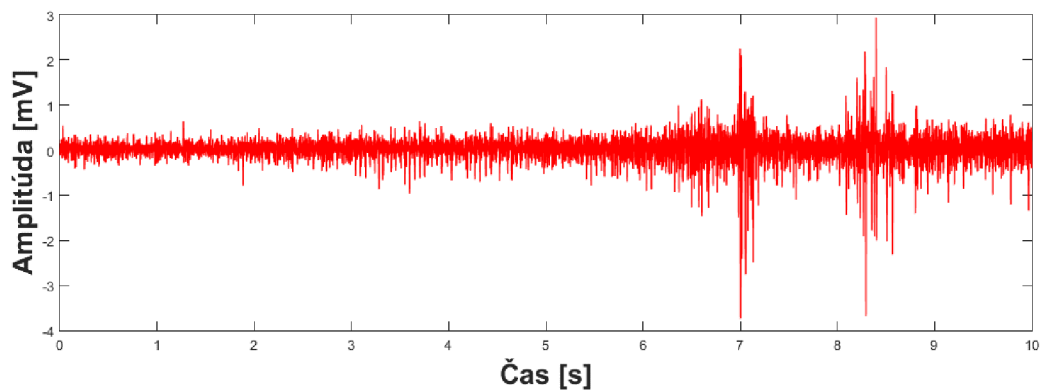


*Obrázok 12. 19 : EKG biosignál, koniec nočnej jazdy, 10s*

Signály EMG podobne ako počas dennej jazdy nevykazujú veľké rozdiely v amplitúdach v originálnej oblasti. Bližšie informácie o tom, či bola identifikovaná únava u vodiča na konci nočnej jazdy, budú analyzované a následne overené v nasledujúcej kapitole.



*Obrázok 12. 20 : EMG biosignál, začiatok nočnej jazdy, 10s.*



*Obrázok 12. 21 : EMG biosignál, koniec nočnej jazdy, 10s.*

## 13. OVERENIE POMOCOU POPISNEJ ŠTATISTIKY

Štatistické metódy umožňujú jednotlivé biologické dáta medzi sebou porovnávať. Popisná štatistika je považovaná za najzákladnejšiu štatistickú metódu, ktorá umožňuje prvotný náhľad na súbor dát. Jej úlohou je sumarizovať informácie, spracovávať ich vo forme grafov a tabuliek a vypočítavať ich číselné charakteristiky polohy, medzi ktoré patria aritmetický priemer, modus, medián, minimum, maximum a charakteristiky variability, medzi ktoré patria, smerodajná odchýlka, rozptyl a ďalšie.

Aritmetický priemer je definovaný ako podiel súčtu a počtu vstupných dát, modus predstavuje najpočetnejšiu hodnotu súboru dát a medián prostrednú hodnotu usporiadaných dát do neklesajúcej postupnosti. Minimum a maximum sú charakterizované ako prvá a posledná hodnota vo vzostupne usporiadanej množine dát. Smerodajná odchýlka sa získa ako druhá odmocnina z rozptylu hodnôt a rozptyl odpovedá strednej kvadratickej odchýlke jednotlivých nameraných hodnôt od výberového priemeru [27].

Pri analýze spracovaných dát bola použitá časť popisnej štatistiky pre charakteristiku polohy a variability a to konkrétne aritmetický priemer (stredná hodnota), minimum, maximum, smerodajná odchýlka a rozptyl.

Pre overenie, či dochádza k zmenám v príznakoch extrahovaných z biologických signálov bola uskutočnená popisná štatistika. Boli do nej zaradení 4 subjekty, ktorí absolvovali denné aj nočné jazdy, a boli použité významné úseky ich jazd. Subjekty boli muži približne v rovnakom veku.

Pre vyhodnotenie parametrov z EKG boli extrahované príznaky MNN, SDNN a RMSSD, pre EMG energia signálu - Energy a pre EDA jej Stredná hodnota.

### 13.1 Stres

- **Vyhodnotenie parametrov z EKG**

Z extrahovaných úsekov jazdy každého subjektu boli vypočítané len 3 príznaky, a to MNN, SDNN a RMSSD a ich stredná hodnota, smerodajná odchýlka a rozptyl, viď TabuľkyPríloha.pdf. Následne bola vypočítaná celková štatistika pre všetky subjekty a výsledné hodnoty sú ukázané v Tabuľke 13.1. a 13.2. Ako bolo ukázané v [9], s nástupom stresu dochádza k zmene dĺžky intervalu NN približne o 50 ms, zvýšeniu tepu srdca a nárastu hodnôt SDNN a RMSSD. V tabuľke 13.2 je viditeľnejší rozdiel v týchto hodnotách, ku ktorému došlo pri zmene vyvolanej stresovým faktorom. Rozdiel medzi priemernou dĺžkou NN v pokoji a v strese je viac ako 50 ms, stredná hodnota SDNN poklesla v pokojnom stave oproti stresovému na 35% a RMSSD pokleslo zo stresového stavu na pokojný o 48%. Tabuľka 13.1 ukazuje výsledok celkovej štatistiky pre denný záznam, tabuľka 13.2 výsledok celkovej štatistiky pre nočný záznam.



Tabuľka 13. 1 : Hodnoty parametrov analýzy EKG od všetkých subjektov z denného záznamu. Vysvetlivky : SH- stredná hodnota, SO – smerodajná odchýlka, R – rozptyl.

	SDNN - SH [ms]	SDNN - SO [ms]	SDNN - R [ms]	RMSSD - SH [ms]	RMSSD - SO [ms]	RMSSD - R [ms]	NN [ms]
<b>STRES</b>	63,850	66,531	4426,363	44,612	54,891	3013,027	644,107
<b>POKOJ</b>	16,195	6,149	37,808	12,614	6,915	47,819	668,469

Tabuľka 13. 2 : Hodnoty parametrov analýzy EKG od všetkých subjektov z nočného záznamu. Vysvetlivky : SH- stredná hodnota, SO – smerodajná odchýlka, R – rozptyl.

	SDNN - SH [ms]	SDNN - SO [ms]	SDNN - R [ms]	RMSSD - SH [ms]	RMSSD - SO [ms]	RMSSD - R [ms]	NN [ms]
<b>STRES</b>	79,435	41,966	1761,113	59,640	30,583	935,323	671,014
<b>POKOJ</b>	27,724	9,613	92,409	29,973	17,089	292,044	723,236

- **Vyhodnotenie parametrov z EMG**

Z celkových príznakov vypočítaných z EMG signálu bola použitá energia signálu EMG získaná pre pokojné úseky a úseky, u ktorých bol predikovaný nástup stresu. Bola vypočítaná jej stredná hodnota, minimálna hodnota, maximálna hodnota, smerodajná odchýlka a rozptyl. Stresový činiteľ zapríčiňuje zvýšenie energie, čo je demonštrované v tabuľkách 13.3 a 13.4, ktoré sa skladajú z celkovej štatistiky všetkých subjektov. Rozdiely v stredných hodnotách energie signálu EMG u pokojného a stresového stavu sú väčšie v nočnom zázname, kedy pokojný stav vykazoval pokles približne o 96% energie signálu, pričom v dennom zázname došlo k poklesu energie zo stresového stavu do pokojného o necelých 51%. Rozdiely sú viditeľné aj v smerodajnej odchýlke a rozptylu hodnôt.

Tabuľka 13. 3 : Hodnoty parametrov analýzy EMG od všetkých subjektov z denného záznamu. Vysvetlivky: mV<sup>2</sup> – jednotka energie signálu.

	Stredná hodnota [mV <sup>2</sup> ]	Minimum [mV <sup>2</sup> ]	Maximum [mV <sup>2</sup> ]	Smerodajná odchýlka [mV <sup>2</sup> ]	Rozptyl [mV <sup>2</sup> ]
<b>STRES</b>	0,2136	0,0262	0,4272	0,1804	0,0325
<b>POKOJ</b>	0,1061	0,0170	0,2910	0,1286	0,0165

Tabuľka 13. 4 : Hodnoty parametrov analýzy EMG od všetkých subjektov z nočného záznamu. Vysvetlivky: mV<sup>2</sup> – jednotka energie signálu.

	<b>Stredná hodnota [mV<sup>2</sup>]</b>	<b>Minimum [mV<sup>2</sup>]</b>	<b>Maximum [mV<sup>2</sup>]</b>	<b>Smerodajná odchýlka [mV<sup>2</sup>]</b>	<b>Rozptyl [mV<sup>2</sup>]</b>
<b>STRES</b>	1,8596	0,5029	3,1676	1,3146	1,7282
<b>POKOJ</b>	0,0768	0,0394	0,0957	0,0253	0,0006

- **Vyhodnotenie parametrov z EDA**

Hlavným ukazovateľom nástupu stresu u EDA je stredná hodnota, podľa ktorej boli dopočítané minimum, maximum, smerodajná odchýlka a rozptyl. Záznamy EDA boli vo veľkom počte prípadov nepoužiteľné, alebo boli príliš ovplyvnené prekrývajúcimi sa QRS komplexmi. Z tohto dôvodu bola výsledná štatistika v značnej miere skreslená. Podobne ako u EKG a EMG najväčšie rozdiely vznikli u nočného záznamu, viď tabuľka 13.6, kedy maximálna hodnota stresu dosiahla 0,7149 S a stredná hodnota 0,6081 S. V dennom zázname, viď 13.5, sú rozdiely medzi stresom a pokojom nepatrné, maximum a stredná hodnota boli dosiahnuté v oboch prípadoch takmer rovnaké.

Tabuľka 13. 5 : Hodnoty parametrov analýzy EDA od všetkých subjektov z denného záznamu.

	<b>Stredná hodnota [S]</b>	<b>Minimum [S]</b>	<b>Maximum [S]</b>	<b>Smerodajná odchýlka [S]</b>	<b>Rozptyl [S]</b>
<b>STRES</b>	0,3676	0,2309	0,5042	0,1933	0,0373
<b>POKOJ</b>	0,3604	0,2163	0,5044	0,2037	0,0415

Tabuľka 13. 6 : Hodnoty parametrov analýzy EDA od všetkých subjektov z nočného záznamu.

	<b>Stredná hodnota [S]</b>	<b>Minimum [S]</b>	<b>Maximum [S]</b>	<b>Smerodajná odchýlka [S]</b>	<b>Rozptyl [S]</b>
<b>STRES</b>	0,6081	0,4928	0,7149	0,1113	0,0124
<b>POKOJ</b>	0,3690	0,2478	0,5227	0,1403	0,0197

## 13.2 Únava

- **Vyhodnotenie parametrov z EKG**

Tabuľka 13.7. znázorňuje výsledky celkovej štatistiky subjektov, kedy boli priemerované hodnoty NN, SDNN a RMSSD každého subjektu zo záznamu začiatku jazdy a tesne pred ukončením jazdy v tom istom úseku. Stredná hodnota NN narástla na konci jazdy o 49 ms a poklesla hodnota SDNN a RMSSD, čo podľa [8] znamená nástup únavy. Tabuľka 13.8. ukazuje výsledky štatistiky z nočnej jazdy, kedy bol zaznamenaný úsek na začiatku diaľnice a chvíľu predtým, čo vodič použil zjazd z diaľnice. V tomto prípade očakávaný nárast strednej hodnoty NN nenastal, došlo k zníženiu o necelých 7 ms. Značný vplyv na túto výchylku mal prvý subjekt, u ktorého došlo k značnému nárastu MNN, SDNN a RMSSD na konci jazdy, vid' TabuľkyPRÍLOHA.pdf.

Tabuľka 13. 7 : Hodnoty parametrov analýzy EKG zo začiatku jazdy a pred ukončením dennej jazdy pre všetky subjekty.

	<b>SDNN - SH [ms]</b>	<b>SDNN - SO [ms]</b>	<b>SDNN - R [ms]</b>	<b>RMSSD - SH [ms]</b>	<b>RMSSD - SO [ms]</b>	<b>RMSSD - R [ms]</b>	<b>NN [ms]</b>
<b>Začiatok jazdy</b>	39,023	52,270	2732,102	42,452	60,292	3635,145	647,690
<b>Koniec jazdy</b>	24,063	16,736	280,100	13,776	7,023	49,329	696,536

Tabuľka 13. 8 : Hodnoty parametrov analýzy EKG zo začiatku diaľnice a pred zjazdom z diaľnice počas nočnej jazdy pre všetky subjekty.

	<b>SDNN - SH [ms]</b>	<b>SDNN - SO [ms]</b>	<b>SDNN - R [ms]</b>	<b>RMSSD - SH [ms]</b>	<b>RMSSD - SO [ms]</b>	<b>RMSSD - R [ms]</b>	<b>NN [ms]</b>
<b>Začiatok diaľnice</b>	26,813	14,292	204,270	26,341	14,354	206,023	736,103
<b>Koniec diaľnice</b>	16,208	3,533	12,479	20,892	6,852	46,947	729,248

- **Vyhodnotenie parametrov z EMG**

Tabuľka 13.9. ukazuje výsledky štatistiky zo záznamu začiatku jazdy a chvíľu pred ukončením jazdy. Predpokladalo sa, že na konci jazdy budú subjekty vykazovať známky únavy prejavene vo zvýšení energie signálu. Ani u jedného subjektu tento predpoklad nebol potvrdený, práveže došlo k poklesu energie signálu. Tabuľka 13.10. obsahuje

hodnoty výslednej štatistiky z úseku jazdy na začiatku diaľnice a chvíľu pred zjazdom z diaľnice. V tomto prípade došlo k značnému nárastu energie signálu na konci jazdy u všetkých subjektov, takže predpoklad o vzniku únavy na konci jazdy bol potvrdený.

Tabuľka 13. 9 : Hodnoty parametrov analýzy EMG zo začiatku jazdy a pred ukončením dennej jazdy pre všetky subjekty. Vysvetlivky: mV<sup>2</sup> – jednotka energie signálu.

	<b>Stredná hodnota [mV<sup>2</sup>]</b>	<b>Minimum [mV<sup>2</sup>]</b>	<b>Maximum [mV<sup>2</sup>]</b>	<b>Smerodajná odchýlka [mV<sup>2</sup>]</b>	<b>Rozptyl [mV<sup>2</sup>]</b>
<b>Začiatok jazdy</b>	0,4468	0,0977	1,2954	0,5716	0,3267
<b>Koniec jazdy</b>	0,1444	0,0219	0,4576	0,2097	0,0440

Tabuľka 13. 10 : Hodnoty parametrov analýzy EMG zo začiatku diaľnice a pred zjazdom z diaľnice počas nočnej jazdy pre všetky subjekty. Vysvetlivky: mV<sup>2</sup> – jednotka energie signálu.

	<b>Stredná hodnota [mV<sup>2</sup>]</b>	<b>Minimum [mV<sup>2</sup>]</b>	<b>Maximum [mV<sup>2</sup>]</b>	<b>Smerodajná odchýlka [mV<sup>2</sup>]</b>	<b>Rozptyl [mV<sup>2</sup>]</b>
<b>Začiatok jazdy</b>	0,0947	0,0172	0,1745	0,0870	0,0076
<b>Koniec jazdy</b>	0,1946	0,0856	0,3486	0,1307	0,0171

- **Vyhodnotenie parametrov z EDA**

Výrazný pokles strednej hodnoty signálu EDA je spojený s nástupom únavy [28], ale tento prípad nebol preukázaný ani u jedného subjektu. V prípade dennej jazdy, tabuľka 13.11., a v prípade nočnej jazdy, tabuľka 13.12., došlo naopak k nárastu strednej hodnoty EDA. Denná jazda vykazovala menšie rozdiely v stredných hodnotách na začiatku a na konci jazdy.

Tabuľka 13. 11 : Hodnoty parametrov analýzy EDA zo začiatku jazdy a pred ukončením dennej jazdy pre všetky subjekty.

	<b>Stredná hodnota [S]</b>	<b>Minimum [S]</b>	<b>Maximum [S]</b>	<b>Smerodajná odchýlka [S]</b>	<b>Rozptyl [S]</b>
<b>Začiatok jazdy</b>	0,5339	0,1442	0,9546	0,4061	0,1649
<b>Koniec jazdy</b>	0,5852	0,2734	0,9781	0,3593	0,1291

Tabuľka 13. 12 : Hodnoty parametrov analýzy EDA zo začiatku diaľnice a pred zjazdom z diaľnice počas nočnej jazdy pre všetky subjekty.

	<b>Stredná hodnota [S]</b>	<b>Minimum [S]</b>	<b>Maximum [S]</b>	<b>Smerodajná odchýlka [S]</b>	<b>Rozptyl [S]</b>
<b>Začiatok jazdy</b>	0,2661	0,0392	0,4931	0,3209	0,1030
<b>Koniec jazdy</b>	0,5183	0,4929	0,5436	0,0358	0,0013

## 14. NÁVRH KLASIFIKAČNÝCH MODELOV A OPTIMALIZÁCIA HYPERPARAMETROV

Do klasifikačných modelov vstupovali príznaky, ktoré boli vypočítané z úsekov 18 jázd, ktoré absolvovali 9 subjekty počas dennej a nočnej hodiny. Subjekty boli mužského pohlavia, približne v rovnakom veku. Pre detekciu stresu a únavy bolo použitých 5 klasifikačných algoritmov strojového učenia, ktorými sú rozhodovacie stromy, DT, klasifikátor Naivný Bayes, NB, metóda podporných vektorov, SVM, k-najbližší sused, KNN a logistická regresia, LR. Princípy jednotlivých algoritmov boli vysvetlené v kapitole 4.

Po tom ako bola aplikovaná vhodná selekčná metóda pre detekciu stresu aj únavy, ktorou bol algoritmus NCA, viď podkapitoly 11.3 a 11.5, ktorého princíp je uvedený v podkapitole 3.3, boli pre vstup do klasifikačných modelov pre detekciu stresu vybrané príznaky, pre pripomenutie sú uvedené v tabuľke 14.1 na prvom riadku a príznaky vybrané pre vstup do klasifikačných modelov pre únavu sú v druhom riadku. Počet meraní (pozorovaní) bolo po selekcii 168 pre stres a pre únavu 117.

Vytvorený bol vektor Target, ktorý pozostával z logických hodnôt a jeho dĺžka bola rovná počtu všetkých pozorovaní (meraní) pre stres a únav zvlášť. Miestu, ktorému prislúchal index riadku merania, ktoré vykazovalo kľudový úsek alebo úsek bez únavy bola pridelené logická 0 a zvyšným miestam logická 1, ktorá vypovedala o stresovom a únavovom úseku.

Následne boli vytvorené indexy pre k-násobnú krížovú validáciu podľa počtu meraní a so selekčným parametrom 10. Model bol trénovaný 10 krát, pričom dáta boli podľa indexov rozdelené na trénovacie a testovacie. Na optimalizáciu hyperparametrov bol použitý algoritmus Vyhľadávacia mriežka, keďže poskytoval lepšie výsledky oproti náhodnému prehľadávaniu a Bayesovskej optimalizácii. Základné informácie sú uvedené v kapitole 6. V nasledujúcich podkapitolách budú zhrnuté návrhy klasifikačných metód a optimalizácia hyperparametrov pre detekciu stresu a únavy súčasne.

Tabuľka 14. 1 : Vybrané príznaky pre vstup do klasifikačných algoritmov.

<b>STRESS</b>	Energy	RMS	Variance	Power Signal	minNN	maxNN	MNN	MAV
<b>ÚNAVA</b>	minNN	MNN	SDNN	pNN20	TINN	MAV	MDF	HR

### 14.1 Štandardizácia príznakov

Pre modely NB, SVM, KNN a LR je nutné previesť štandardizáciu, aby bolo zabezpečené, že každý príznak má rovnakú váhu. Štandardizácia zvyčajne znamená, že

sa škálujú údaje tak, aby mali strednú hodnotu rovnú 0 a rozptyl rovný 1 [52]. Štandardizácia bola súčasťou príkazu klasifikačného modelu a nasledujúca rovnica popisuje jej výpočet, kde  $\mu$  a  $\sigma$  znamenajú priemer a rozptyl hodnôt v rámci jedného príznaku [53].

$$x' = \frac{x - \mu}{\sigma} \quad (14.1)$$

## 14.2 Rozhodovacie stromy

Táto klasifikačná metóda bola zvolená ako prvá a následne sa práca rozšírila o ďalšie klasifikačné metódy strojového učenia. Základy tohto algoritmu sú uvedené v kapitole 4. Optimalizácia hyperparametrov prebiehala automaticky a bola súčasťou príkazu na vytvorenie modelu rozhodovacieho stromu. Tabuľka 14.2 obsahuje hyperparametre, ktoré je možné optimalizovať pri tejto klasifikačnej metóde, a tie ktoré boli vybrané pre klasifikáciu stresu a únavy.

Parameter `MaxNumSplits` definuje maximálny počet rozhodovacích rozdelení alebo uzlov, čiže počet rozdelení neprekročí túto hodnotu, ale môže byť aj menší. `MinLeafSize` definuje minimálny počet listových uzlov stromu. `SplitCriterion`, alebo kritérium rozdelenia sa používa na zníženie maximálnej odchýlky. Parameter `NumVariablesToSample` určuje počet prediktorov, ktoré sa majú náhodne vybrať pre každé rozdelenie, takže musí byť vždy kladné číslo a môže dosahovať aj maximálny počet prediktorov.

Tabuľka 14. 2 : Tabuľka s vybranými optimalizovanými parametrami pre rozhodovací strom pre stres a únavu.

PARAMETRE	STRES	ÚNAVA
<code>MaxNumSplits</code>	150	104
<code>MinLeafSize</code>	4	1
<code>SplitCriterion</code>	'gdi'	'gdi'
<code>NumVariablesToSample</code>	'all'	'all'

## 14.3 Klasifikátor Naivný Bayes

Tento klasifikátor dosahuje výborných výsledkov podľa viacerých štúdií a preto bol vybraný aj pre túto prácu. Princíp tohto klasifikačného algoritmu je uvedený v kapitole 4. Tabuľka 14.3 popisuje výber optimalizačných hyperparametrov pre klasifikáciu stresu a únavy.

Parameter `DistributionNames` používa klasifikátor na modelovanie údajov, pričom obsahuje možnosti ako sú 'kernel', 'multinomiál', a 'normal'. `Width` parameter určuje šírku vyhladzovacieho jadra. `Kernel` definuje vyhladzovací typ jadra, ako 'epanechnikov', 'normal', atď. `Support` slúži ako podpora hustoty vyhladenia jadra, pričom 'unbounded' znamená, že sa jedná o skutočné hodnoty.

Tabuľka 14. 3 : Tabuľka s vybranými optimalizovanými parametrami pre klasifikátor Naivný Bayes pre stres a únavu.

PARAMETRE	STRES	ÚNAVA
DistributionNames	'kernel'	'kernel'
Width	0,0019	5,08e-12
Kernel	'normal'	'normal'
Support	'unbounded'	'unbounded'

## 14.4 Metoda podporných vektorov

Základné informácie o tejto metode boli vysvetlené v kapitole číslo 4. V nasledujúcej tabuľke sú vypísané optimálne parametre, ktoré vstupovali do klasifikačného modelu podporných vektorov.

`BoxConstraint` je parameter, ktorý definuje obmedzenie rámčeka a môže nadobúdať rôznych kladných hodnôt. `KernelScale` parameter je určený ako parameter mierky jadra, a slúži nato, že všetky prvky predikčnej matice sú touto hodnotou podelené a následne sa použijú na výpočet Gramovej matice. `KernelFunction` sa používa na výpočet prvkov Gramovej matice. `PolynomialOrder` prezentuje poradie funkcií polynomiálneho jadra. `Standardize` parameter sa používa na štandardizáciu prediktorov pred vstupom do klasifikačného modelu.

Tabuľka 14. 4 : Tabuľka s vybranými optimalizovanými parametrami pre klasifikačnú metódu podporných vektorov pre stres a únavu.

PARAMETRE	STRES	ÚNAVA
BoxConstraint	10 000	0,0215
KernelScale	0,4642	1,00e-03
KernelFunction	'linear'	'linear'
PolynomialOrder	-	-
Standardize	'true'	'true'



## 14.5 K – najbližších susedov

Základné informácie o tejto klasifikačnej metóde sú popísané v kapitole 4 a v tabuľke 14.5 sa nachádzajú optimálne parametre pre túto metódu.

Parameter Distance označuje použitú metriku na meranie vzdialeností. DistanceWeight je váhovanou funkciou vzdialeností a môže byť ‚equal‘, ‚inverse‘, atď. Exponent je definovaný ako Minkowského exponent vzdialeností. Počet najbližších susedov, ktorý sa má hľadať pri predikcii. Standardize parameter sa používa na štandardizáciu prediktorov pred vstupom do klasifikačného modelu.

Tabuľka 14. 5: Tabuľka s vybranými optimalizovanými parametrami pre klasifikačnú metódu K-najbližší sused pre stres a únavu.

PARAMETRE	STRES	ÚNAVA
Distance	‚mahalanobis‘	‚correlation‘
DistanceWeight	‚equal‘	‚equal‘
Exponent	-	-
NumNeighbors	1	9
Standardize	‚true‘	‚true‘

## 14.6 Logistická regresia

Princíp logistickej regresie je vysvetlený v kapitole 4. Pri tomto modeli nedochádzalo k automatickej optimalizácii hyperparametrov, ale niektoré parametre boli nastavené manuálne, podľa najlepšej dosiahnutej presnosti a AUC hodnoty a sú uvedené v tabuľke 14.6.

Distribution parameter nazývaný distribúcia môže nadobúdať hodnoty ‚normal‘, ‚binomial‘, ‚poisson‘, ‚gamma‘ a ‚inverse gaussian‘. Parameter Link je tzv. Funkcia spojenia, ktorá sa má použiť namiesto všeobecnej funkcie spojenia.

Tabuľka 14. 6 : Tabuľka s vybranými parametrami pre klasifikačnú metódu logistická regresia pre stres a únavu.

PARAMETRE	STRES	ÚNAVA
Distribution	‚binomial‘	‚binomial‘
Link	‚logit‘	‚logit‘

## 15. VÝSLEDKY KLASIFIKAČNÝCH MODELOV

Každý klasifikačný model sa testoval na trénovacej sade 10 krát a následne sa výsledné hodnoty úspešnosti modelu v podobe AUC spriemerovali. Tento proces bol vykonaný z dôvodu, že pri k-násobnej krížovej validácii sa vyberá vždy náhodná trénovacia a testovacia množina.

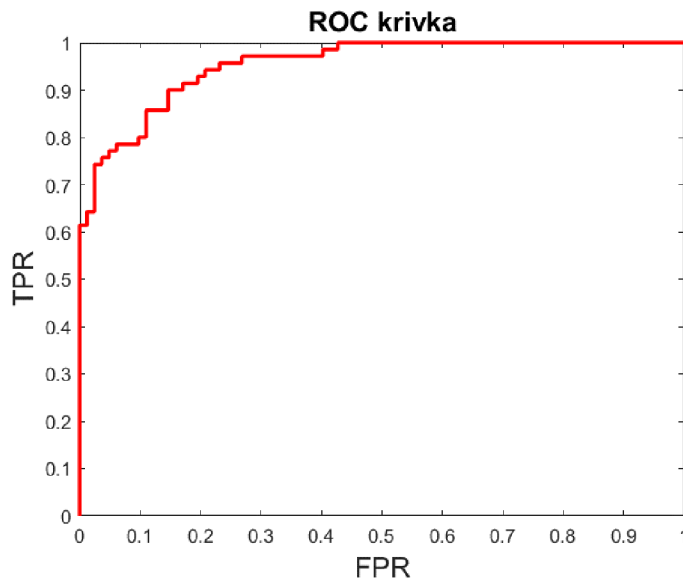
U vytvorených modeloch sa hodnotila klasifikačná presnosť ACC, oblasť pod krivkou AUC, senzitivita TPR a špecificita FPR.

### 15.1 Hodnotenie klasifikačných modelov pre stres

V tabuľke 15.1 sú konečné spriemerované hodnoty úspešnosti klasifikačných modelov pre detekciu stresu. Ako spoľahlivý indikátor úspešnosti natrénovaného modelu je považovaná hodnota AUC, vid' tretí stĺpec, kde je vyznačená najvyššia hodnota, ktorá bola dosiahnutá použitím klasifikačnej metódy Naivný Bayes. Druhú najvyššiu hodnotu AUC dosiahol rozhodovací strom a najnižšiu hodnotu dosiahla metóda Logistická regresia. Na grafe 15.1 je zobrazená ROC krivka klasifikačného modelu Naivný Bayes. Klasifikačná presnosť sa pohybuje u klasifikátorov v rozmedzí od 69,6 % - 81,4%, pričom najlepšia presnosť bola dosiahnutá pri metóde Logistickej regresie. Najlepšia senzitivita bola dosiahnutá pri metóde Naivný Bayes a špecificita TNR u Logistickej regresii. Celková klasifikačná úspešnosť dosahovala pomerne vysoké hodnoty aj napriek tomu, že niektoré biosignály neboli vhodné pre spracovanie a samotných dát bol menší počet.

Tabuľka 15. 1 : Výsledné zhodnotenie úspešnosti klasifikačných modelov pre stres.

MODEL	ACC [%]	AUC [%]	TPR	TNR
<b>Rozhodovací strom</b>	72	92,2	0,714	0,727
<b>Naivný Bayes</b>	77,4	93,98	0,846	0,688
<b>Metóda podporných vektorov</b>	72	83	0,780	0,649
<b>K-Najbližší sused</b>	69,6	88,2	0,758	0,623
<b>Logistická regresia</b>	81,4	78,8	0,757	0,864



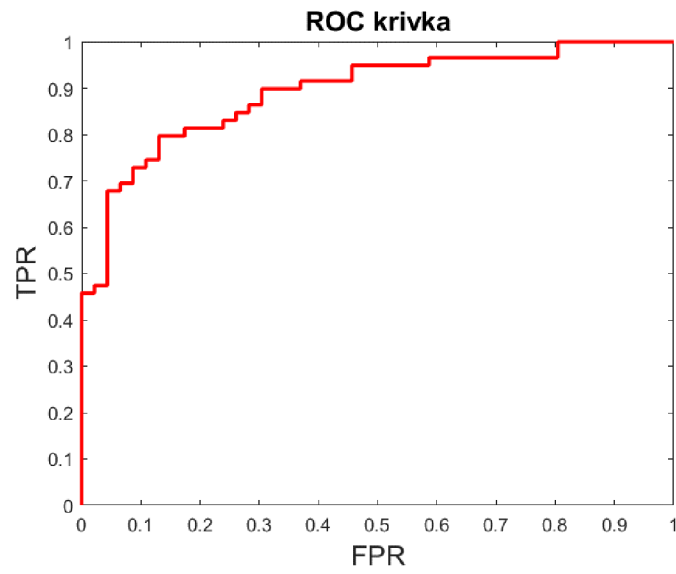
Graf 15. 1 : ROC krivka ako výsledok hodnotenia klasifikačného modelu Naivný Bayes pre stres.

## 15.2 Hodnotenie klasifikačných modelov pre únavu

Úspešnosť klasifikačných modelov pre únavu dosahuje pomerne slabých výsledkov. Klasifikačná presnosť neprekročila hranicu 69,5 %, čo je stále veľmi malá úspešnosť. Problém vytvárali dáta, ktoré neboli veľmi konzistentné a ďalším faktorom bol subjektívny výber úsekov, kde sa nejavili známky únavy a únavové úseky, napr. monotónna jazda po diaľnici. AUC hodnota prekvapivo dosiahla najvyššie číslo 92,8 %, čo pôsobí skoro nerealisticky. Vzhľadom nato, že klasifikačná presnosť v tomto prípade dosiahla hodnotu 50,4 %, tak sa nedá táto hodnota považovať za relevantnú. Naproti tomu rozhodovací strom poskytol hodnotu AUC 84,1 % a zároveň klasifikačnú úspešnosť 60,7 %, čo je druhá najvyššia hodnota ACC v tabuľke 15.2. V grafe 15.2 bola vykreslená ROC krivka metódy Naivný Bayes.

Tabuľka 15. 2 : Výsledné zhodnotenie úspešnosti klasifikačných modelov pre únavu.

MODEL	ACC [%]	AUC [%]	TPR	TNR
<b>Rozhodovací strom</b>	60,7	84,1	0,451	0,727
<b>Naivný Bayes</b>	50,4	92,8	0,157	0,772
<b>Metóda podporných vektorov</b>	56,4	62,1	0,569	0,561
<b>K-Najbližší sused</b>	49,6	79,8	0,255	0,682
<b>Logistická regresia</b>	69,5	64,4	0,712	0,674



*Graf 15. 2 : ROC krivka ako výsledok hodnotenia klasifikačného modelu Naivný Bayes pre únavu.*

## 16. TESTOVANIE NA REÁLNYCH DÁTACH

Pre overenie úspešnosti klasifikácie stresu a únavy boli klasifikačné modely vytvorené v tejto práci, použité na detekciu stresu a únavy v reálnych dátach. Zo súboru „.csv“, ktorý obsahoval záznamy biologických signálov, boli vybrané signály EKG a EMG. Následne boli signály rozdelené na intervaly dlhé 10 sekúnd, resp. 10 000 vzoriek podľa  $f_{vz} = 1000$  Hz. Dané intervaly signálov boli upravené podľa transformačnej funkcie, došlo k predspracovaniu signálov a extrakcii príznakov. Príznaky boli upravené do vhodnej formy pre správnu predikciu. Nasledovala predikcia triedy každého 10 sekundového záznamu všetkými klasifikačnými modelmi a výsledok bol uložený do tabuľky, kde počet riadkov odpovedal počtu 10 sekundových záznamov a každý stĺpec odpovedal predikčnej hodnote jedného z klasifikačných modelov, viď príklad uvedený v tabuľke 16.1.

### 16.1 Predikcia stresu

Pre predikciu stresu bol vybraný denný hodinový záznam jedného zo subjektov. Tabuľka popisuje 50 sekundový záznam, čiže každý riadok vystihuje 10 sekundový záznam a každý prvok tabuľky definuje predikovanú triedu pre daný klasifikátor. Riadok označený číslom 2 je úsek jazdy, kedy došlo k stresu u vodiča. Záznam biosignálov EKG a EMG z toho úseku je na obrázku 16.2 a 16.3. Fotka z videokamery z daného úseku jazdy je na obrázku 16.1, kde je vidieť že u vodiča došlo k stresu kvôli chodcovi, ktorý vstúpil na vozovku z rady áut.

Ako najlepší model sa v tomto prípade opäť javí metóda Podporných vektorov, ktorá vyhodnotila daný úsek ako stresový, pričom mu predchádzal kľudový úsek. Ďalšou metódou, ktorá zareagovala podobne bola Logistická regresia, ale stres u vodiča bol predikovaný aj naďalej, pričom u metódy Podporných vektorov bol stres predikovaný ďalších 10 sekúnd a následne bol predikovaný kľud.

Rozhodovací strom predikoval obidve triedy počas hodinového záznamu a ako je vidieť v tabuľke 16.1 predikoval stres a následne predikoval kľud, čo je možné považovať za správnu klasifikáciu stresu. Klasifikátor Naivný Bayes v tomto 50 sekundovom zázname predikoval stres, ale v iné okamihy jazdy, ktoré vykazovali dlhodobé kľudové úseky, napríklad jazdu po prázdnej d'iaľnici, predikoval hodnotu 0 – kľud. Metóda K-najbližších susedov predikovala tiež stres a následne kľud, ale nedá sa to považovať za relevantný výsledok, keďže celú minútu pred týmto záznamom bol predikovaný stres.

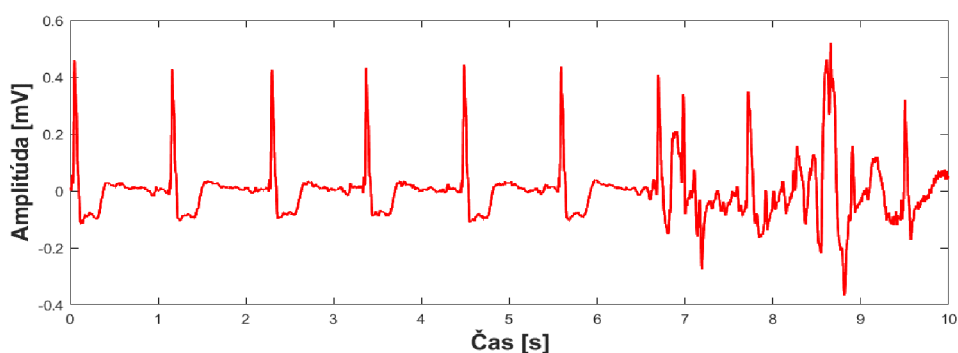
Metóda Podporných vektorov v častiach záznamu, u ktorých jednoznačne došlo k stresu, napr. chodec náhle vkročil do vozovky, auto pred nami prudko zabrzdilo a lopta sa skotúlala na cesty, vykazovala predikčnú triedu stres a úsek pred udalosťou a po tejto udalosti bol predikovaný ako kľudový.

Tabuľka 16. 1 : Predikovaná trieda v 50 sekundovom zázname pre každý klasifikátor.

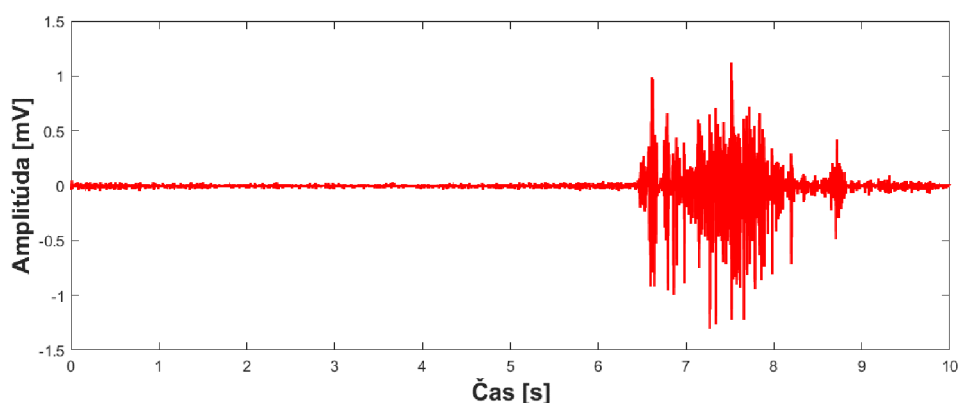
	Rozhodovací strom	Naivný Bayes	Podporné vektory	K-Najbližších susedov	Logistická regresia
1.	,stres´	,stres´	,kl'ud´	,stres´	,kl'ud´
2.	,stres´	,stres´	,stres´	,stres´	,stres´
3.	,stres´	,stres´	,stres´	,stres´	,stres´
4.	,kl'ud´	,stres´	,kl'ud´	,kl'ud´	,stres´
5.	,kl'ud´	,stres´	,stres´	,stres´	,stres´



Obrázok 16. 1 : Úsek jazdy vykazujúci stres u vodiča – chodec vstúpil na vozovku z rady áut.



Obrázok 16. 2 : Záznam biosignálu EKG, u ktorého bol predikovaný stres.



Obrázok 16. 3 : Záznam biosignálu EMG, u ktorého bol predikovaný stres.

## 16.2 Predikcia únavy

Keďže klasifikátory únavy dosahovali nie veľmi veľkých presnosti, najlepší 68,6 %, tak aj predikcia na reálnych dátach nebola až taká úspešná. Predikcia bola uskutočnená na hodinovom nočnom zázname toho istého subjektu, u ktorého bola vykonaná predikcia na dennom zázname. Klasifikátor Naivný Bayes vykazoval po celý záznam triedu 1-únavu, klasifikátor K-Najbližší sused po celú dobu 0 – bez únavy, čo potvrdzuje fakt, že tieto klasifikačné modely dosahovali najnižšiu klasifikačnú presnosť. Metóda Podporných vektorov a Logistická regresia vykazovala približne rovnaké triedy až na zopár výnimiek, ale väčšinu času obe klasifikátory predikovali bez únavový stav. Po videní záznamu v dané momenty, kedy bola predikovaná únava sa šofér nachádzal na diaľnici. Klasifikátor Rozhodovací strom predikoval triedu 1 – únava častejšie než Podporné vektory a Logistická regresia, ale pri videní záznamu z videokamery sa zistilo, že predikčné hodnoty mali pravdepodobne občas náhodný charakter.

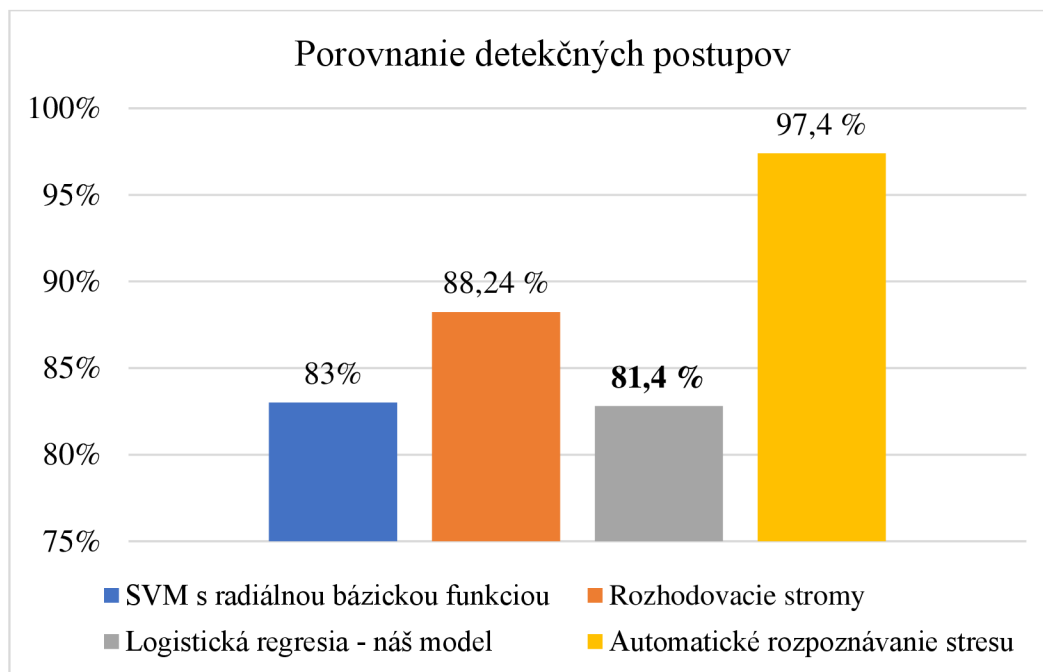
## 17. POROVNANIE S PUBLIKÁCIAMI

Problematikou detekcie stresu a únavy sa zaoberá veľké množstvo publikácií, ktoré popisujú rôzne metódy na docielenie adekvátnych výsledkov a sú uvedené v tabuľkách 17.1 a 17.2. Najviac používanými technikami sú metódy strojového učenia, najmä metóda podporných vektorov SVM, ktorá dosahuje vysokú presnosť.

Z klasifikačných algoritmov, ktoré boli otestované v tejto práci, dosahoval najvyššiu klasifikačnú presnosť pre detekciu stresu klasifikátor Logistická regresia – 81,4 %. Dosiahol len o 1,6 % horšiu klasifikačnú presnosť ako 2. publikovaná metóda uvedená v tabuľke 17.1. V stĺpcovom grafe 17.1 sú znázornené výsledky z publikácií a hodnoty najlepšieho klasifikátora v tejto práci.

Tabuľka 17. 1 : Publikované metódy detekcie stresu.

Metóda	Klasifikačná presnosť [%]	Odkaz
Automatické rozpoznávanie stresu	97,4 %	<a href="#">[11]</a>
SVM s radiálnou bázickou funkciou	83 %	<a href="#">[56]</a>
Rozhodovacie stromy	88,24%	<a href="#">[57]</a>



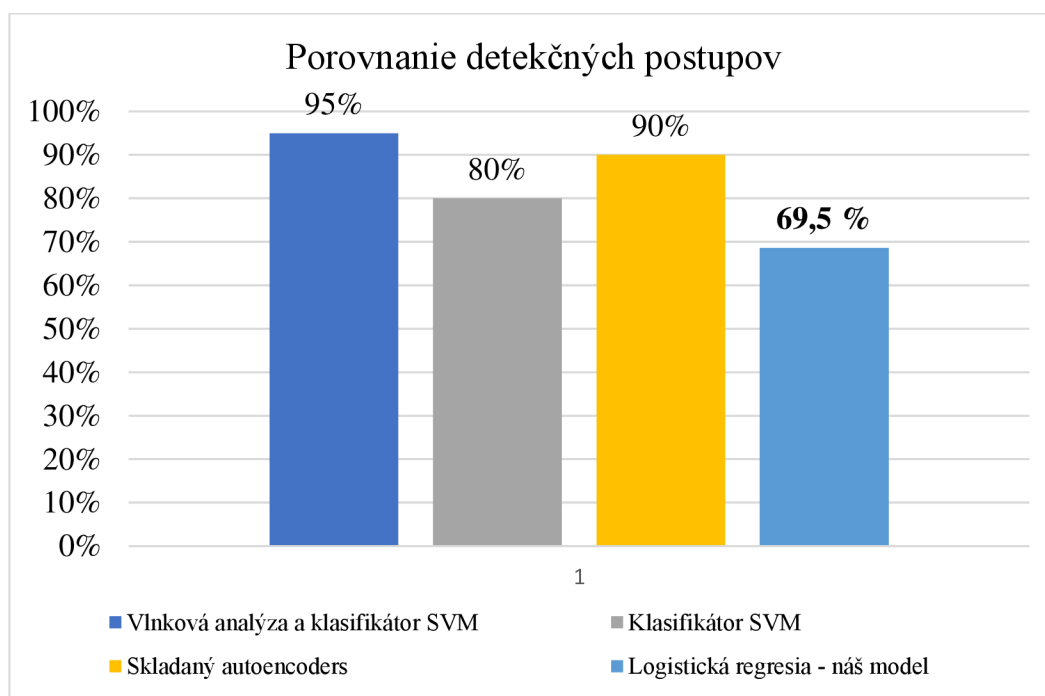
Graf 17. 1 : Porovnaná klasifikačná presnosť publikovaných detekčných algoritmov pre stres s najlepším klasifikačným modelom tejto práce.



V tejto práci bolo vytvorených 5 klasifikačných modelov pre detekciu únavy a najvyššiu klasifikačnú presnosť dosahoval opäť klasifikátor Logistická regresia – 69,5 %. Oproti úspešnosti algoritmov v publikovaných článkoch je naša úspešnosť pomerne nízka. Veľký vplyv na výslednú úspešnosť mal subjektívny výber úsekov jász, ktoré u vodiča vykazovali únavu a úseky, kde únava nebola preukázaná. V stĺpcovom grafe 17.2 sú znázornené výsledky z publikácií a hodnoty najlepšieho klasifikátora v tejto práci.

Tabuľka 17. 2 : Publikované metódy detekcie únavy.

Metóda	Klasifikačná presnosť [%]	Odkaz
Vlnková analýza variability srdcového rytmu a klasifikátor SVM	95%	<a href="#">[7]</a>
Klasifikátor SVM	80%	<a href="#">[55]</a>
Skladaný autoencoders	90 %	<a href="#">[58]</a>



Graf 17. 2 : Porovnaná klasifikačná presnosť publikovaných detekčných algoritmov pre únavu s najlepším klasifikačným modelom tejto práce.

## DISKUSIA

Problematika detekcie stresu a únavy u vodiča zohráva dôležitú úlohu pri zvyšujúcej sa frekvencii áut na cestách, monotónnych cestách, ktorými sú diaľnice a pracovnej vyčerpanosti vodiča. Existuje rada automobiliek, ktoré využívajú vo svojich autách bezpečnostné systémy na upozornenie vodiča pred nástupom únavy a blížiacou sa hrozbe.

Hlavným zdrojom dát pre detekciu stresu a únavy v tejto práci boli biologické signály EKG a EMG, ktoré boli vybrané z databázy reálnych záznamov z jász na automobile BMW rady 5. Ektodermálna aktivita bola z detekcie vylúčená kvôli nesprávnym a poškodeným záznamom. Z 18 denných a nočných jász, ktoré absolvovali 9 vodiči boli vybrané 10 sekundové úseky biosignálov, u ktorých sa dalo predpokladať, že vodič vykazoval známky únavy, stresu, alebo kľudné úseky a úseky bez únavy. Výber týchto dát sa uskutočnil len podľa subjektívneho názoru tvorca diplomovej práce, pretože neboli dostupné žiadne záznamy, ktoré by výber relevantných dát potvrdili. Výsledná presnosť detekcie je len taká dobrá, ako bol subjektívny výber dát spoľahlivý. Do úvahy je potrebné zahrnúť fakt, že každý subjekt, ktorý sa zúčastnil merania je unikátny, reaguje na podnety rôzne a biosignály môžu byť ovplyvnené samotnými fyziologickými parametrami daného subjektu. Príkladom boli subjekty, u ktorých nebola počas celej jazdy spozorovaná zmena v biosignáloch a zapríčinili teda veľkú variabilitu vo vybraných biosignáloch pre klasifikáciu stavov vodiča. Iným príkladom bol subjekt, ktorý mal celú jazdu rapídne zvýšenú tepovú frekvenciu. Ďalším faktorom sú vonkajšie vplyvy, ktoré nebolo možné ovplyvniť a boli rôzne pre jednotlivé subjekty. Príkladom je frekvencia výskytu áut, práca na ceste, poveternostné podmienky a spolusediaci v aute, ktorí môžu pozitívne, alebo aj negatívne pôsobiť na vodiča.

Táto práca sa venovala klasifikácii daných stavov vodiča do 4 skupín, a to kľud, stres, únava a bezúnavový stav. Detekcia stresu oproti únave dosahovala vyšších hodnôt a to z dôvodu väčšieho počtu signálov, ktorých príznaky boli použité ako vstup do klasifikačných modelov a vyššej pravdepodobnosti spoľahlivého výberu úsekov jász, v ktorých došlo k prudkému brzdeniu auta idúceho pred subjektom, lopty kotúľajúcej sa cez cestu a chodcom, ktorí vkročili na vozovku spoza rady áut. Najvyššia dosiahnutá úspešnosť správnej klasifikácie stresu a kľudu bola 81,4 % a pri porovnaní s dostupnými publikáciami bolo zistené, že náš najlepší model dosiahol len o 1,6 % nižšiu úspešnosť oproti publikovanej metóde. V publikáciách prevládajú metódy strojového učenia, najmä klasifikátor Podporných vektorov. Vyššiu úspešnosť by mohli priniesť nové dáta z hodiniek Empatica, ktoré mali byť súčasťou práce, ale z dôvodu nekonzistentnosti nameraných dát boli z práce vylúčené. Vyššiu úspešnosť by mohol zabezpečiť aj výber úsekov s menším počtom sekúnd a pomocný zápis spolusediaceho, ktorý by zaznamenal čas, kedy by vodič potvrdil, že mohlo s veľkou pravdepodobnosťou dôjsť k fyziologickej zmene.

Klasifikácia únavových a bez únavových stavov dosahovala pri najlepšom modeli klasifikačnú presnosť 69,5 %. Oproti uvedeným publikáciám je nami dosiahnutá úspešnosť klasifikácie pomerne nízka. Ako už bolo spomenuté, veľký vplyv na konečnú úspešnosť mal subjektívny výber úsekov jász a veľkú rolu zohrávali aj monotónne úseky diaľnice v nočných záznamoch subjektov, ktoré mohli a nemuseli vykazovať známky únavy. Z tohto dôvodu by pomohli vo výbere opäť pomocné záznamy, ktoré by obsahovali časové údaje, kedy došlo u vodiča k nástupu únavy.

Klasifikačná úspešnosť sa pridávaním nových dát neustále zvyšovala, čo je znakom spoľahlivého klasifikátora a zároveň to znamená, že nové namerané dáta by mohli prispieť k zvýšeniu presnosti klasifikačných modelov. K novým dátam by mohli byť pridané aj ďalšie spoľahlivé biologické signály, napríklad ektodermálna aktivita, záznam respirácie a signál z eyetrackeru, ktoré by mohli pomôcť k správnej detekcii stresu a únavy.

## ZÁVER

Diplomová práca sa venuje klasifikácii stavov vodiča na základe získaných relevantných príznakov jednotlivých biologických signálov z vybraných úsekov jász. Je zameraná na výber takých príznakov, ktoré najviac vystihujú zmenu biosignálov od normálne stavu k stresovému stavu, alebo k stavu, ktorý vykazuje únavu a podporujú metodiku klasifikačných modelov na vybrané záznamy jász.

Táto diplomová práca využíva klasifikačné modely, vďaka ktorým je docielená detekcia stresu a únavy. Práca je rozdelená do sedemnástich kapitol, ktoré reprezentujú body zadania.

Prvých šesť kapitol pozostáva z teoretických znalostí, ktoré sú potrebné pre vypracovanie danej témy. Na začiatku je vypracovaná rešerš dostupných publikácií, ktoré sa zaoberajú detekciou stresu a únavy z biologických signálov u vodiča. Následne je uvedené spracovanie biosignálov, extrakcia a selekcia príznakov, princípy klasifikačných algoritmov, hodnotenie klasifikačných modelov a nakoniec optimalizácia hyperparametrov.

V ďalšej časti sa uskutočňuje zoznámenie s databázou jász, ktorá bola vytvorená v spolupráci s Ústavom soudního inženýrství v Brne a boli vybrané vhodné úseky jász vďaka vytvorenému softvéru v Pythone na Ústave biomedicínského inženýrství, ktorých príklady sú v práci uvedené.

Následne je uvedené predspracovanie biologických signálov, ktoré prebiehalo v prostredí Matlab, výber relevantných príznakov, eliminácia nevhodných dát a selekcia príznakov. Vybrané príznaky boli overené pomocou popisnej štatistiky.

Hlavnou úlohou diplomovej práce bola detekcia únavy a stresu, ktorá prebiehala na základe klasifikačných modelov strojového učenia, typu Rozhodovací strom, Naivný Bayes, metóda Podporných vektorov, K-Najbližších susedov a Logistická regresia. Jednotlivé modely boli vylepšené vďaka optimalizácii hyperparametrov pomocou metódy Vyhľadávacej mriežky.

Detekcia stresu dosahovala uspokojivé výsledky v podobe 77,1 % klasifikačnej presnosti pre metódu Podporných vektorov, ktorá je najpoužívanejšou technikou pri detekcii stresu v publikovaných metódach. Naproti tomu detekcia únavy dosahovala najvyššiu klasifikačnú presnosť len 58,1 % pre klasifikátor Rozhodovací strom. Táto nízka presnosť bola zapríčinená subjektívnym výberom úsekov vykazujúcich únavu alebo úsekov bez únavy a veľkou variabilitou dát, ktorú spôsobovali subjekty, u ktorých nedochádzalo počas jazdy k veľkým rozdielom v príznakoch.

Vytvorené modely boli použité na predikciu stavu vodiča na reálnych dátach. Jednalo sa o hodinové denné a nočné záznamy. Stres bol predikovaný na dennom zázname, kde pomerne uspokojivé výsledky boli zistené pre metódu Podporných vektorov a trochu menej spoľahlivých výsledkov bolo docielených pomocou metódy Logistická regresia. Na nočnom zázname, u ktorého boli hľadané úseky s únavou,

poskytovali klasifikačné modely neuspokojivé výsledky, pretože po celý záznam predikovala väčšina z modelov stav únavový a ostatné klasifikátory poskytovali skôr náhodné výsledky.

Posledný bod práce obsahuje vypracovaný prehľad dostupných publikácií, ktoré sa zaoberajú tou istou problematikou, a bolo vykonané porovnanie ich klasifikačných presností s presnosťou, ktorá bola dosiahnutá v tejto práci. Najviac využívanou technikou v publikáciách bola metóda Podporných vektorov, keďže poskytuje najlepšie výsledky.

## LITERATÚRA

- [1] MAY, J., C. BALDWIN. *Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies*. (2009). *Transportation Research Part F: Traffic Psychology and Behaviour*. 12. 218-224. doi: 10.1016/j.trf.2008.11.005.
- [2] THIFFAULT, P., J. BERGERON. *Monotony of road environment and driver fatigue: a simulator study*. *Accid Anal Prev*. 2003 May;35(3):381-91. doi: 10.1016/s0001-4575(02)00014-3. PMID: 12643955.
- [3] NOLEN-HOEKSEMA, S., B. L. FREDRICKSON, G. R. LOFTUS, A. WILLEM, WAGENAAR. *Psychologie Atkinsonové a Hilgarda*. Vydání třetí, přepracované. Přeložil Hana ANTONÍNOVÁ. Praha: Portál, 2012. ISBN 978-80-262-0083-3.
- [4] RIGAS, G., K. CHRISTOS, P. BOUGIA, D. FOTIADIS. *A reasoning-based framework for car driver's stress prediction*. (2008). 627 - 632. doi: 10.1109/MED.2008.4602162.
- [5] SIKANDER, G., S. ANWAR. "Driver Fatigue Detection Systems: A Review," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339-2352, June 2019, doi: 10.1109/TITS.2018.2868499.
- [6] JUNG, S., H. SHIN, W. CHUNG. "Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel," in *IET Intelligent Transport Systems*, vol. 8, no. 1, pp. 43-50, Feb. 2014, doi: 10.1049/iet-its.2012.0032.
- [7] LI, G., Chung W. Y. *Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier*. (2013). *Sensors* (Basel, Switzerland). 13. 16494-16511. doi: 10.3390/s131216494
- [8] LEE, H. B., CHOI J. M., KIM J. S., KIM Y. S., BAEK H. J., RYU M. S., SOHN R. H., PARK K. S. *Nonintrusive biosignal measurement system in a vehicle*. (2007). In 29th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society, *EMBC'07* (pp. 2303-2306). [4352786] (Annual International Conference of the IEEE Engineering in Medicine and Biology -Proceedings).

- [9] RASTGOO, M.N. *Driver Stress Level Detection Based on Multimodal Measurements*. Ph.D. Thesis, Queensland University of Technology, Queensland, Australia, 2019. Available online: <https://eprints.qut.edu.au/134144/>
- [10] SOLAZ, J., HERNÁNDEZ J. L., BANDE D., RODRÍGUEZ N., VELLE S., GERPE J., MEDINA E. *Drowsiness Detection Based on the Analysis of Breathing Rate Obtained from Real-time Image Recognition*. (2016). *Transportation Research Procedia*. 14. 3867-3876. doi: 10.1016/j.trpro.2016.05.472
- [11] HEALEY, J. A., PICARD R. W. "Detecting stress during real-world driving tasks using physiological sensors," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156-166, June 2005, doi: 10.1109/TITS.2005.848368.
- [12] MILAGRO, J., GIL E., GARZÓN-REY J. M., AGUILO J., BAILÓN R., "Inspiration and expiration dynamics in acute emotional stress assessment," *2017 Computing in Cardiology (CinC)*, Rennes, 2017, pp. 1-4, doi: 10.22489/CinC.2017.261-411.
- [13] What is a Notch Filter? In: *EverythingRF* [online]. Aug 29, 2018 [cit. 2020-11-19]. Dostupné z: <https://www.everythingrf.com/community/what-is-a-notch-filter>.
- [14] BALASUBRAMANIAN, V., ADALARASU K. *EMG-based analysis of change in muscle activity during simulated driving*. (2007). *Journal of Bodywork and Movement Therapies*. 11. 151-158. doi: 10.1016/j.jbmt.2006.12.005.
- [15] SASIDHARAN, D., VENUGOPAL G., "Analysis of Surface EMG Signals under Fatigue Conditions using Sum of Sines Models," *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2020, pp. 0849-0852, doi: 10.1109/ICCSP48568.2020.9182049.
- [16] SOWMYA, N., S. SRIVARSHINI, N. SHANMATHI, R. MENAKA. "Stress Diagonisis Using EMG Signals," *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, 2018, pp. 1-4, doi: 10.1109/ICCTCT.2018.8551028.
- [17] KARABULUT, D., F. ORTES, Y. Z. ARSLAN, M. A. ADLI, "Comparative evaluation of EMG signal features for myoelectric controlled human arm prosthetics", *Biocybern. Biomed. Eng.*, vol. 37, no. 2, pp. 326-335, 2017.

- [18] TOMASZEWSKI, J., T.G. AMARAL, O.P. DIAS, A. WOLCZOWSKI, M. KURZYNSKI. *EMG signal classification using neural network with AR model coefficients*. (2009). 318-325. doi: 10.3182/20090819-3-PL-3002.00056.
- [19] GOKGOZ, E., A. SUBASI. *Comparison of decision tree algorithms for EMG signal classification using DWT*. (2015). Biomedical Signal Processing and Control. 18. doi. 10.1016/j.bspc.2014.12.005.
- [20] ZHONGLIANG, Y., Y. CHEN, J. WANG, H. GONG. *Recognizing the breathing resistances of wearing respirators from respiratory and sEMG signals with artificial neural networks*. (2017). International Journal of Industrial Ergonomics. 58. 47-54. doi: 10.1016/j.ergon.2017.02.001.
- [21] SUGANTHI, L., M. S. S. DEVI, M. ILAKKIYA, G. TAMIZHARASI, "Effect of physical stress on human electro dermal activity," *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai, 2017, pp. 2505-2508, doi: 10.1109/ICECDS.2017.8389903.
- [22] AMALAN, S., A. SHYAM, A. S. ANUSHA. "Electrodermal Activity based Classification of Induced Stress in a Controlled Setting," *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Rome, 2018, pp. 1-6, doi: 10.1109/MeMeA.2018.8438703.
- [23] HSIEH, C., Y. CHEN, W. K. BEH, A. A. WU. "Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection," *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, Nanjing, China, 2019, pp. 330-335, doi: 10.1109/SiPS47522.2019.9020321.
- [24] ŠIMOŇÁKOVÁ, S. *Systémy pro akvizici fyziologických dat řidiče při jízdě*. Brno, 2019. Dostupne take z: <https://www.vutbr.cz/studenti/zav-prace/detail/118335>. Bakalářska práce. Vysoké učení technické v Brne, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce Martin Mézl.
- [25] KOZUMPLÍK, J. *Analýza biologických signálů* [online]. FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ, Brno, květen 2012 [cit. 2021-4-6].



- [26] CICIORA, W., J. FARMER, M. ADAMS. Chapter 4 - *Digital Modulation* [online]. (2004). [cit. 2020-12-31], doi: 10.1016/B978-1-55860-828-3.X5000-0.
- [27] PROCHÁZKA, B. *Biostatistika pro lékaře: principy základních metod a jejich interpretace s využitím statistického systému R*. V Praze: Univerzita Karlova v Praze, nakladatelství Karolinum, 2015. ISBN 978-80-246-2782-3.
- [28] MALATHI, D., J. D. D. JAYASEELI, S. MADHURI, K. SENTHILKUMAR. *Electrodermal Activity Based Wearable Device for Drowsy Drivers*. (2018). Journal of Physics: Conference Series. 1000. 012048. doi: 10.1088/1742-6596/1000/1/012048.
- [29] KAMINSKI, B., M. JAKUBCZYK, P. SZUFEL. *A framework for sensitivity analysis of decision trees*. (2018). *Central European journal of operations research*, 26(1), 135–159, doi: 10.1007/s10100-017-0479-6.
- [30] *Ergoneers* [online]. [cit. 2021-01-04]. Dostupné z: [https://www.ergoneers.com/en/hardware/dikablis-glasses/?fbclid=IwAR0ahIA4P1sJXDdTUTk7-LGx-CAO06EVcQ8dWK9F9ZOtcuVdj\\_Q76ed-Lc0](https://www.ergoneers.com/en/hardware/dikablis-glasses/?fbclid=IwAR0ahIA4P1sJXDdTUTk7-LGx-CAO06EVcQ8dWK9F9ZOtcuVdj_Q76ed-Lc0)
- [31] LEONEL, J. *Classification Methods in Machine Learning* [online]. Oct 9, 2019 [cit. 2021-5-4]. Dostupné z: <https://medium.com/@jorgesleonel/classification-methods-in-machine-learning-58ce63173db8>
- [32] SIDANA, M. *Intro to types of classification algorithms in Machine Learning* [online]. Feb 28, 2017 [cit. 2021-5-4]. Dostupné z: <https://medium.com/sifium/machine-learning-types-of-classification-9497bd4f2e14>
- [33] GANDHI, R. *Support Vector Machine — Introduction to Machine Learning Algorithms* [online]. Jun 7, 2018 [cit. 2021-5-4]. Dostupné z: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [34] SARMIENTO, D. L. R. *Digital Processing of Electromyographic Signals for Control: MATLAB Central File Exchange* [online]. 18 Jul, 2018 [cit. 2021-5-5]. Dostupné z: <https://www.mathworks.com/matlabcentral/fileexchange/68245-digital-processing-of-electromyographic-signals-for-control>

- [35] BENEDEK, M., CH. KAERNBACH. *Decomposition of skin conductance data by means of nonnegative deconvolution* [online]. Jul 2010 [cit. 2021-5-5]. Dostupné z: doi:10.1111/j.1469-8986.2009.00972.x
- [36] NOTO, T., G. ZHOU, S. SCHUELE, J. TEMPLER a C. ZELANO. *Automated analysis of breathing waveforms using BreathMetrics: a respiratory signal processing toolbox* [online]. 07 July 2018, , Pages 583–597 [cit. 2021-5-6]. Dostupné z: doi:<https://doi.org/10.1093/chemse/bjy045>
- [37] GUPTA, T. *Performance Measures for Classification Models: And methods to evaluate the performance of a classifier* [online]. Dec 3, 2020 [cit. 2021-5-6]. Dostupné z: <https://towardsdatascience.com/performance-measures-for-classification-models-a486c8976bf1>
- [38] KOLAR, R. *Introduction: Machine Learning* [online]. Department of Biomedical Engineering, FEEC, Brno University of Technology, 2020 [cit. 2021-5-6].
- [39] MCCORMICK, Ch. *K-fold cross-validation, with Matlab code: Computer Vision and Machine Learning Projects and Tutorials* [online]. July 31, 2013 [cit. 2021-5-6]. Dostupné z: <https://chrisjmccormick.wordpress.com/2013/07/31/k-fold-cross-validation-with-matlab-code/>
- [40] NARKHEDE, S. *Understanding AUC - ROC Curve* [online]. Jun 26, 2018 [cit. 2021-5-6]. Dostupné z: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
- [41] KOLAR, R. *Dimensionality reduction: Machine Learning* [online]. Department of Biomedical Engineering, FEEC, Brno University of Technology, 2020 [cit. 2021-5-11].
- [42] GLEN, S.. *Statistics How To: Chi-Square Statistic: How to Calculate It / Distribution* [online]. [cit. 2021-5-11]. Dostupné z: <https://www.statisticshowto.com/probability-and-statistics/chi-square/>
- [43] BROWNLEE, J. *Difference Between Classification and Regression in Machine Learning*. Start Machine Learning [online]. December 11, 2017 [cit. 2021-5-12]. Dostupné z: <https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/>

- [44] IPPOLITO, P. P. *Hyperparameters Optimization: An introduction on how to fine-tune Machine and Deep Learning models using techniques such as: Random Search, Automated Hyperparameter Tuning and Artificial Neural Networks Tuning*. [online]. Sep 26, 2019 [cit. 2021-5-12]. Dostupné z: <https://towardsdatascience.com/hyperparameters-optimization-526348bb8e2d>
- [45] *Why Is Random Search Better Than Grid Search For Machine Learning* [online]. 14/06/2018 [cit. 2021-5-12]. Dostupné z: <https://analyticsindiamag.com/why-is-random-search-better-than-grid-search-for-machine-learning/>
- [46] BISSUEL, A. *Hyper-parameter optimization algorithms: a short review* [online]. Apr 16, 2019 [cit. 2021-5-12]. Dostupné z: <https://medium.com/criteo-engineering/hyper-parameter-optimization-algorithms-2fe447525903>
- [47] *Electrocardiography (ECG) Sensor Datasheet* [online]. In: . 2020, s. 3 [cit. 2021-5-12]. Dostupné z: [https://biosignalsplux.com/downloads/docs/datasheets/Electrocardiography\\_\(ECG\)\\_Datasheet.pdf](https://biosignalsplux.com/downloads/docs/datasheets/Electrocardiography_(ECG)_Datasheet.pdf)
- [48] SHAFFER, F., J. P. GINSBERG. *An Overview of Heart Rate Variability Metrics and Norms* [online]. 2017 Sep 28 [cit. 2021-5-12]. Dostupné z: doi:10.3389/fpubh.2017.00258
- [49] *ECG Basics: Determining Rate* [online]. [cit. 2021-5-12]. Dostupné z: <https://www.healio.com/cardiology/learn-the-heart/ecg-review/ecg-interpretation-tutorial/determining-heart-rate>
- [50] POURMOHAMMADI, S., A. MALEKI. *Computer Methods and Programs in Biomedicine: Stress detection using ECG and EMG signals: A comprehensive study* [online]. September 2020 [cit. 2021-5-12]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0169260719320528#bib0036>
- [51] *MathWorks: Band power* [online]. [cit. 2021-5-12]. Dostupné z: <https://www.mathworks.com/help/signal/ref/bandpower.html>
- [52] GELLER, S. *Normalization vs Standardization — Quantitative analysis* [online]. Apr 4, 2019 [cit. 2021-5-14]. Dostupné z:

<https://towardsdatascience.com/normalization-vs-standardization-quantitative-analysis-a91e8a79cebf>

- [53] GRUS, J. *Data Science from Scratch, 2nd Edition*. May 2019. O'Reilly Media. ISBN 9781492041139.
- [54] GOLDBERGER, J., S. ROWEL, G. HINTON a R. SALAKHUTDINOV. *Neighbourhood Components Analysis*. Department of Computer Science, University of Toronto [online]. [cit. 2021-5-15]. Dostupné z: <https://www.cs.toronto.edu/~hinton/absps/nca.pdf>
- [55] AWAIS, M., N. BADRUDDIN a M. DRIEBERG. *Driver Stress Level Detection Based on Multimodal Measurements* [online]. 2017 Aug 31 [cit. 2021-5-16]. Dostupné z: doi:10.3390/s17091991
- [56] MUNLA, N., M. KHALIL, A. SHAHIN a A. MOURAD. *Driver Stress Level Detection Using HRV Analysis*. Lebanese University Tripoli, Lebanon [online]. 12 November 2015, , 4 [cit. 2021-5-16]. Dostupné z: doi:10.1109/ICABME.2015.7323251
- [57] KESHAN, N., P. V. PARIMI a I. BICHINDARITZ. *Machine learning for stress detection from ECG signals in automobile drivers* [online]. 28 December 2015 [cit. 2021-5-16]. Dostupné z: <https://ieeexplore.ieee.org/document/7364066>
- [58] BHARDWAJ, R., P. NATRAJAN a V. BALASUBRAMANIAN. *Study to Determine the Effectiveness of Deep Learning Classifiers for ECG Based Driver Fatigue Classification* [online]. 27 May 2019 [cit. 2021-5-16]. Dostupné z: doi:10.1109/ICIINFS.2018.8721391

## ZOZNAM SYMBOLOV A SKRATIEK

### Skratky:

EKG	Elektrokardiografia
PPG	Fotopletysmografia
HRV	Variabilita srdcovej frekvencie
EMG	Elektromyografia
EDA	Elektrodermálna aktivita
GPS	Globálny lokalizačný systém
HRV	Variabilita srdcového rytmu
EOG	Elektrookulografia
ANS	Autonómny nervový systém
PSO	Algoritmus roja častíc
FFT	Fourierová transformácia
DWT	Diskrétna vlnková transformácia
STFT	Krátkodobá fourierová transformácia
PCA	Analýza hlavných komponentov
LDA	Lineárna diskriminačná analýza
DT	Rozhodovacie stromy
NB	Naivný Bayes
SVM	Metóda Podporných vektorov
KNN	K-Najbližších susedov
LR	Logistická regresia
RF	Náhodný les
ROC	Krivka prevádzkovej charakteristiky
AUC	Oblasť pod krivkou
MRMR	Algoritmus Maximálnej relevancie a minimálnej nadbytočnosti
NCA	Analýza susedských komponent
SO	Smerodajná odchýlka
R	Rozptyl
Napr.	Napríklad
A pod.	A podobne
Tzv.	Takzvané

### Symboly:

Hz	Hertz
Ms	milisekundy
mV <sup>2</sup>	Jednotka energie signálu

mV	milivolt
S	Siemens
s	sekundy
„.csv“	Prípona súboru excel
F <sub>vz</sub>	Vzorkovacia frekvencia

## ZOZNAM PRÍLOH

- **SÚBOR „TABULKYPRÍLOHA“** – ROZŠÍRENÉ TABUĽKY S POPISNOU ŠTATISTIKOU
- **SÚBOR „STRES“** – OBSAHUJE HLAVNÝ SKRIPT „STRES.M“; SKRIPT NA TESTOVANIE „TESTING.M“; SKRIPT NA VÝPOČET POPISNEJ ŠTATISTIKY; FUNKCIE „HRVANALYSIS.M“ A „EMGANALYSIS.M“; SÚBORY S EXTRAHOVANÝMI PRÍZNAKMI PRE KLASIFIKAČNÉ MODELY, ČASŤ HODINOVÉHO ZÁZNAMU PRE TESTOVANIE KLASIFIKÁTOROV NA REÁLNYCH DÁTACH A PRÍKLADY VYBRANÝCH ÚSEKOV JÁZD PRE POPISNÚ ŠTATISTIKU.
- **SÚBOR „ÚNAVA“** - OBSAHUJE HLAVNÝ SKRIPT „STRES.M“; SKRIPT NA TESTOVANIE „TESTING.M“; FUNKCIE „HRVANALYSIS.M“ A „EMGANALYSIS.M“; SÚBORY S EXTRAHOVANÝMI PRÍZNAKMI PRE KLASIFIKAČNÉ MODELY A ČASŤ HODINOVÉHO ZÁZNAMU PRE TESTOVANIE KLASIFIKÁTOROV NA REÁLNYCH DÁTACH
- ODKAZ NA PROGRAM V PYTHONĚ: <https://github.com/branislavhesko/D-LAB-visualizer>