



**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**ELEKTROS IR ELEKTRONIKOS FAKULTETAS**

**Gediminas Žylius**

**INTELEKTUALIOS VAIRAVIMO ELGSENOS ĮVERTINIMO  
SISTEMOS KŪRIMAS IR TYRIMAS**

Baigiamasis magistro projektas

**Vadovas**  
Doc. dr. Vygandas Vaitkus

**KAUNAS, 2015**

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**ELEKTROS IR ELEKTRONIKOS FAKULTETAS**  
**AUTOMATIKOS KATEDRA**

**INTELEKTUALIOS VAIRAVIMO ELGSENOS ĮVERTINIMO  
SISTEMOS KŪRIMAS IR TYRIMAS**

Baigiamasis magistro projektas  
Valdymo technologijos (621H66001)

**Vadovas**  
Doc. dr. Vygandas Vaitkus

**Recenzentas**

**Projektą atliko**  
Gediminas Žylius

**KAUNAS, 2015**



Žylius, G. Intelektualios vairavimo elgsenos įvertinimo sistemos kūrimas ir tyrimas. *Magistro laipsnio* baigiamasis projektas / vadovas doc. dr. Vygandas Vaitkus; Kauno technologijos universitetas, elektros ir elektronikos fakultetas, elektros energetikos sistemų katedra.

Kaunas, 2015. 58 psl.

## SANTRAUKA

*Vairavimo elgsena gali būti nusakoma kaip individualūs vairuotojo gebėjimai valdyti tam tikros rūšies transporto priemonę. Šie bruožai formuoja vairavimo stilių, kuris priklausomai nuo tyrimo srities gali būti išskaidomas į: agresyvų ir saugų vairavimą, ekonomišką ir neekonomišką vairavimą, blaivų ir apsvaigusį vairavimą ir t.t. Šiuolaikinėms transporto kompanijoms naudinga analizuoti savo darbuotojo (ar kliento) vairavimo elgseną, nes nuo to priklauso transporto priemonės tarnavimo laikas, gedimų intensyvumas, tiekiamų transporto paslaugų kokybė. Agresyviu vairavimo pobūdžiu vadinamos visos neigiamos vairavimo savybės dėl kurių gadinama transporto priemonė ir padidinama rizika sukelti eismo įvykį ir pablogėja transporto paslaugų kokybė.*

*Šiame darbe atliekami vairavimo elgsenos tyrimai panaudojant inercinių vairavimo signalų – trijų ašių  $G$  – jėgos jutiklio (akselerometro) informaciją. Tyrimo objektai yra du: agresyvaus ir saugaus vairavimo stiliaus (elgsenos) atpažinimo tyrimas kada  $G$  – jutiklio signalų informacija surenkama iš lengvosios transporto priemonės kuri vairuojama arba agresyviai arba saugiai; vairuotojo elgsenos identifikavimo (vairuotojų tarpusavio klasifikavimo) tyrimas, kurio metu surenkami signalai iš viešojo transporto autobuso, kai jis vairuojamas dviejų vairuotojų.*

*Atlikta analizė ir sudaromi algoritmai pirminio signalo apdorojimo srityje: triukšmo lygio sumažinimas ir išskirčių pašalinimas kombinuojant slankios medianos ir slankaus vidurkio filtrus, taip pat stovėjimo duomenų pašalinimas (viešojo transporto autobuso signaluose) taikant Gauso mikštūrų modelius. Suformuoti požymių išgavimų metodai iš apdorotų signalų tiek dažnių, tiek laiko srityje. Laiko srities požymiai: signalų histogramų kvantiliai; atstumai tarp kvantilių; koreliacijos koeficientai tarp signalų; signalų duomenų ribos viršijimo intensyvumai; trūkių (signalų pirmos eilės skirtumų) standartiniai nuokrypiai. Dažnių srityje išgaunami požymiai: santykinis galios spektro intensyvumas įvairiose dažnių juostose. Atliekamas agresyvaus ir saugaus vairavimo klasifikavimo tyrimas naudojant atraminių vektorių, neuroninių tinklų, atsitiktinių miškų,  $k$ -artimiausių kaimynų klasifikavimo algoritmus. Atliekamas požymių atrinkimo ir transformavimo metodų efektyvumo įvertinimas agresyvaus ir saugaus vairavimo klasifikavimo uždaviniui spręsti. Agresyvaus ir saugaus vairavimo stiliaus rezultatai rodo, jog galima pasiekti ~95% klasifikavimo tikslumą naudojant 10 vairavimo požymių iš visų 78 (didensį nei naudojant visus požymius), bei naudojant tik  $X$  ašies signalo 4 geriausius požymius galima pasiekti ~91,5% klasifikavimo tikslumą. Viešojo transporto autobuso vairuotojo elgsenos atpažinimo uždaviniui spręsti panaudojami tie patys 78 požymiai, atliekamas požymių atrinkimas naudojant informaciją apie požymio priklausymą vienam ar kitam vairuotojui. Atrinkus požymius atliekamas vairuotojų klasifikavimo tikslumo įvertinimas. Nustatyta, jog naudojant visus 78 požymius pasiekiamas ~79,4% vairuotojų klasifikavimo tikslumas, naudojant 20 pagal svarbumą rangotų požymių pasiekiamas ~79,2% klasifikavimo tikslumas, o naudojant 7 geriausius vairavimo požymius, kai 6 iš jų išgaunami iš  $X$  ašies signalų duomenų, pasiekiamas ~77% klasifikavimo tikslumas. Atlikus vairuotojų agresyvaus ir saugaus elgsenos, bei autobusų vairuotojų tarpusavio klasifikavimo tyrimus pastebėta, jog abiem atvejais požymių rinkiniai duodantys didžiausią tikslumą sutampa arba yra panašūs, to pasekoje galima teigti jog vairuotojų elgsenos skirtumai yra pagrinde sąlygoti skirtingo vairavimo agresyvumo laipsnio.*

*Reikšminiai žodžiai: klasifikavimas, požymių išgavimas, vairavimo elgsenos identifikavimas, signalų apdorojimas,  $G$  – jėgos jutiklis.*

Žylius, Gediminas. Development and Research on Intelligent Driving Style Evaluation System. Final project of *master's degree* / supervisor doc. dr. Vygandas Vaitkus; Kaunas University of Technology, Faculty of Electrical and Electronics Engineering, Department of Electrical Power Systems

Kaunas, 2015. 58 psl.

## SUMMARY

*Driving behavior is determined by individual driving capability features to control particular type of vehicle. These features form particular driving style, which depending on application may be categorized as: aggressive and safe driving, eco and fuel-demanding driving, sober and drowsy driving etc. Nowadays it is useful for transport companies to analyze driving behavior of their employees (or clients), because driving behavior affects vehicle life time, vehicle failure intensity, quality of transport services. Aggressive driving are considered all negative driving features that damage vehicle, increase risk of traffic accident, worsen transport services.*

*In this work a detail driving behavior analysis is performed using inertial measurement sensor – 3- axis G-sensor (accelerometer) signal information, which is attached to vehicle. There are to objectives of this research: recognition of aggressive and safe driving behavior when signal information is obtained from vehicle that is driven in aggressive and safe driving styles; driving behavior identification (driver classification) when signals are obtained from public transport bus that is driven by two drivers (one at a time).*

*Signal processing analysis is performed and algorithms formed that reduce noise level and detect outliers in signal by combining moving median and average filters. Also a special signal processing is done on public transport bus signals when stopping time must be extracted and deleted from raw signal. This is performed by applying Gaussian mixture models. Algorithms of feature extraction from accelerometer signal are formed using both time and frequency domain information. From time domain the following features are extracted: quantiles of signal; difference between quantiles; correlation between signals; data threshold violation intensity; standard deviation of jerk signal (first order differences of original signal). In the frequency domain relative power spectral density in particular frequency band are extracted as features for various frequency bands. Experimental analysis of aggressive and safe driving classification performance is investigated when four classification algorithms are used: support vector machines, neural networks, random forest and k-nearest neighbor algorithms. Feature selection using filter methods and feature transformation using principal components is perform in order to study whether this could improve classification accuracy. The results of aggressive and safe driving classification show that using 10 sorted features ~95% accuracy is achieved (that is greater than using all features) and using only 4 features extracted from X axis only ~91.5% accuracy is achieved. When performing public transport driver behavior classification the same all 78 features are extracted from public transport bus accelerometer signals and after feature extraction, a feature selection step is performed again together with driver classification. The results show that when using all 78 features, driving identification accuracy of ~79.4% is achieved, when using best 20 sorted features accuracy of ~79.2% is obtained and by using best 7 features, driver identification rate of ~77% is achieved. 6 of all 7 features are extracted from X axis signal alone. After aggressive and safe driving classification and public transport bus driving behavior classification it was noticed that both feature sets roughly match and are alike. So the conclusion is that differences of driving behavior are conditioned by different level of aggressiveness.*

*Keywords: classification; feature extraction, driving behavior identification; signal processing; G-sensor.*

## TURINYS

ĮVADAS .....	7
1. INERCINIŲ SIGNALŲ PANAUDOJIMAS VAIRAVIMO ELGSENOS ATPAŽINIMUI .....	9
1.1. Išmaniųjų įrenginių kaip jutiklių platformos panaudojimas .....	11
2. AGRESYVIOS IR SAUGIOS VAIRUOTOJO ELGSENOS KLASIFIKAVIMAS .....	13
2.1. Signalų pirminis apdorojimas .....	14
2.2. Požymių išgavimas laiko ir dažnių srityse .....	17
2.3. Agresyvaus ir saugaus vairavimo klasifikavimo metodai .....	23
2.3.1. Atraminių vektorių mašina .....	23
2.3.2. Atsitiktinių miškų klasifikavimo algoritmas .....	25
2.3.3. Neuroninių tinklų klasifikatorius .....	26
2.3.4. k – artimiausių kaimynų klasifikatorius .....	28
2.4. Požymių atrinkimo ir transformavimo metodai .....	28
2.4.1. Porinis Stjudento t – testas .....	29
2.4.2. Santykinė entropija .....	30
2.4.3. Bhattacharyya atstumas .....	30
2.4.4. Wilcoxon rangų sumos testas .....	31
2.4.5. ROC charakteristika .....	32
2.4.6. RELIEFF požymių atrinkimo algoritmas .....	33
2.4.7. Atsitiktinio miško požymių indeksai .....	34
2.4.8. Požymių transformavimas naudojant PCA .....	35
2.5. Klasifikavimo rezultatai .....	36
3. VIEŠOJO TRANSPORTO AUTOBUSŲ VAIRUOTOJŲ ELGSENOS KLASIFIKAVIMAS ..	42
3.1. Signalų apdorojimas .....	42
3.2. Vairavimo elgsenos klasifikavimas .....	48
IŠVADOS .....	52
LITERATŪRA .....	54
PRIEDAI .....	57

## IVADAS

Šiuolaikinėms transporto kompanijoms naudinga analizuoti savo darbuotojo (ar kliento) vairavimo elgseną, nes nuo to priklauso transporto priemonės tarnavimo laikas, gedimų intensyvumas, tiekiamų transporto paslaugų kokybė. Agresyviu vairavimo pobūdžiu vadinamos visos neigiamos vairavimo savybės dėl kurių gadinama transporto priemonė ir padidinama rizika sukelti eismo įvyki. Saugus vairavimo pobūdis priešingas agresyviam. Agresyvus vairavimo pobūdis padidina transporto priemonės išlaikymo kaštus, bei sumažina transporto paslaugų klientų paklausą, priešingai nei saugus vairavimas. Vairavimo elgsenos panašumas į tam tikrą tipinę (etaloninę) gali padėti identifikuoti vairuotojo kvalifikuotumą ar panašumą į tam tikro vairuotojo profilį. Turint sistemą, gebančią įvertinti vairuotojo elgseną būtų galima skatinti saugesnį vairavimą ir keleivių komfortą viešojo transporto sektoriuje, adaptyviau įvertinti vairuotojo rizikingumą ir pagal tai nustatyti draudimo kainą automobilių draudimo kompanijoms, įvertinant vairavimo elgseną nustatyti kas vairuoja transporto priemonę, identifikuoti agresyvų vairavimą ir pavojingus vairavimo manevrus realiaje laike ir pagal tai priimti valdymo sprendimus.

Specializuotų sistemų, skirtų anksčiau minėtai problematikai spręsti rinkoje nėra. O panašaus pobūdžio adaptyvios vairuotoją asistuojančios sistemos (angl. Adaptive Driver Assistance Systems (ADAS)) sistemos yra brangios ir sudėtingai integruojamos dėl daugybės reikalingų įrenginių, be to joms reikalingas realaus laiko komunikacinis tinklas. Senesnėse transporto priemonėse jos gali būti visiškai nepritaikomos. Transporto kompanijoms, pavyzdžiui, susijusioms su maršrutinio transporto paslaugomis, panašios įrangos pritaikymas ne tik per brangus, bet taip pat visos sistemos galimybės nėra išnaudojamos: nereikalingas duomenų apdorojimas realiaje laike, nes duomenų analizė gali būti atlikinėjama, pavyzdžiui, kas mėnesį, nustatinėjant vairuotojo atlyginimą pagal jo vairavimo kokybę per visą laikotarpį.

Šiame darbe tiriama intelektualiai vairavimo elgsenos identifikavimo sistema, kuri gebėtų klasifikuoti vairavimo elgseną pagal požymius išgaunamus iš trijų ašių G – jėgos jutiklio (akselerometro) signalų. Visa darbe pristatoma metodika bus naudojama specializuotos neinvazinės įterptinės sistemos kūrimui. Šios sistemos numatomi privalumai:

- pritaikoma tiek naujoms tiek senoms bet kokios kategorijos transporto priemonėms;
- transporto priemonėje nėra būtinas realaus laiko komunikacinis tinklas (nebent taikoma kartu su realaus laiko valdymo sistema);
- matuojami signalai yra universalūs (matuojama pagreičio sukuriama jėga trijose ortogonaliose ašyse);

- pritaikymas iš vienos transporto priemonės kitai blogiausiu atveju reikalautų minimalaus perkalibravimo (geriausiu atveju perkalibravimo nereiktų);
- duomenų surinkimui reikalingas tik vienas G – jėgos jutiklis su mikroprocesorine įranga, todėl sistema yra labai nebrangi;
- patikimumas (kuo mažiau prietaisų tuo sistema patikimesnė).

Šiame darbe atlikto tyrimo rezultatai rodo, jog su išgautais vairavimo požymiais sistema geba dideliu (~95%) tikslumu klasifikuoti agresyvų ir saugų vairavimą bei identifikuoti dviejų tarpusavyje skirtingų viešojo transporto autobusų vairuotojų vairavimo elgsenas ~79% tikslumu. Abiem atvejais geriausias minimalus požymių rinkinys išgaunamas tik iš išilginės X ašies G – jutiklio (akselerometro) signalo.



## 1. INERCINIŲ SIGNALŲ PANAUDOJIMAS VAIRAVIMO ELGSENOS ATPAŽINIMUI

Inercinių jutiklių (angl. Inertial Measurement Unit (IMU)) pritaikymas intelektualaus transporto sistemose nėra nauja sritis. Per pastaruosius du dešimtmečius buvo atlikta daugybė tyrimų bei eksperimentų, kuriuose šie jutikliai pritaikomi vieni arba kaip dalis sistemos jutiklių.

Jau 1995 m. pasirodžiusiame straipsnyje [1] buvo pasiūlytas giroskopo kartu su greičio jutikliu pritaikymas lengvojoje transporto priemonėje siekiant nustatyti vairuotojo įgūdžius pagal tai koks vairuotojo trajektorijos kreivis lyginant su realiu kelio kreiviu. Kreivis buvo nustatomas kaip atvirksčias dydis trajektorijos apskritimo spinduliui, o spindulys apskaičiuojamas kaip transporto priemonės linijinio (greičio jutiklio duomenų) ir kampinio greičio (giroskopo duomenų) santykis.

2001 m. pasirodė straipsnis [2], kuriame atlikti eksperimentai matuojant keleivio diskomfortą naudojant elektroninį pagreičio jutiklį. Buvo apskaičiuojami išilginių trūkių transporto priemonėje (išilginio pagreičio pirmos išvestinės) intensyvumas tam tikrame siaurame intervale, siekiant sukurti realaus laiko valdymo sistemą, kada valdymo signalas informuoja vairuotoją kad šis modifikuotų savo vairavimo elgesį siekiant išlaikyti keleivių komfortą.

Mitovic [3] savo straipsnyje pasiūlė metodą kaip galima būtų atpažinti vairavimo manevrus iš išilginio ir šoninio pagreičio bei transporto priemonės greičio duomenų naudojant paslėptuosius Markov'o modelius (angl. Hidden Markov Models (HMM)). Pagrindinis trūkumas, jog reikalinga sudaryti manevrų bazę iš istorinių duomenų, kuriais remiantis būtų atliekamas atpažinimas.

Vairuotojų elgesio ir kelio atpažinimo tyrimai [19] taikant HMM metodą buvo atlikti panaudojant dvejetainę metodiką: 1) kelias atpažįstamas apjungiant atpažintus manevrus į visumą; 2) kelias atpažįstamas panaudojant visą kelionės signalą ir jis suskaidomas į tam tikrus manevrus signalo klasteriuose. Eksperimento metu panaudojama informacija surenkama iš transporto priemonės CAN-bus jutiklių tinklo.

Rinkoje egzistuojančių ir tarp draudimo kompanijų populiarių duomenų surinkimo įrenginių (angl. drive recorder) naudojančių GPS ir akselerometro signalus panaudojimas atliktas tyrime [18]. Šiame tyrime naudojant kelio konstrukcijų standartus, nustatančius koku greičiu galima saugiai važiuoti tam tikra kelio posūkio atkarpa nusakoma apskritimo spinduliu bei matuojant kampinio pagreičio bei greičio signalus nustatomas neatsargaus vairavimo kelyje procentinė dalis, nusakanti varuotojo neatsargumą kelyje. Naudojant vairavimo informacijos įrašymo įrenginius (angl. driv recorders) pritaikant jų GPS bei akselerometro jutiklio informaciją buvo atlikti tyrimai tų pačių autorių [23] siekiantys įvertinti vairuotoją, klasifikuojant jo stabdymo pobūdį. Klasterizavimo būdu buvo nustatytos keturios reikšmingos klasės iš istorinių

duomenų, kurių atžvilgiu klasifikuojami nauji vairuotojo stabdymo klasteriai. Tolesni tyrimai [24] vairuotojo rizikai įvertinti naudojant vairavimo informacijos įrašymo įrenginius papildomai naudojo maksimalaus pagreičio kiekvieną minutę bei vidutinio greičio priklausomybes nustatyti vairuotojo pobūdžiui ir rizikai vertinti.

Vairavimo elgsena nustatyti panaudojant GPS ir akcelerometro jutiklius buvo atliktas tyrimas [20], kuriame taikomas šablono atitikimas (angl. pattern matching) klasifikuoti vairavimo stilių. Vairuotojui klasifikuoti panaudojama fuzzy išvadų darymo sistema.

Nešiojama neatsargaus vairavimo atpažinimo sistema buvo pasiūlyta [21]. Ši sistema paprasčiausiai naudoja įtaisą naudojantį dviejų ašių akcelerometrą ir neraiškiają logiką nustatyti blogiems vairavimo įpročiams. Panaši sistema, naudojant dviejų ašių akcelerometrą pasiūlyta tyrime [25]. Jame naudojami išilginiai bei šoniniai pagreičiai apjungiami į vieną vektorių pagal Euler'io formulę ir suformuotas signalas tiriamas naudojant neraiškiają logiką vairavimo stiliui įvertinti.

Vairuotojo elgesio klasifikavimo tyrimai [22] buvo atlikti išilgai kelio atkarpos, tiriant atstumo, greičio bei pagreičio signalus priešais važiuojančios transporto priemonės atžvilgiu. Tyrime buvo panaudojami CAN-bus transporto priemonės signalai bei lazerinis radaras matuoti parametrus priešais esančios transporto priemonės atžvilgiu.

Dviejų ašių akcelerometro signalų skirtumai tarp eksperto bei pradedančio vairuotojo tiriami [28]. Šio tyrimo metu buvo nustatyta jog akcelerometro signalai kiekvienam vairuotojui išlaiko tam tikrus jo individualius požymius, kurie atsikartoja vairavimo toje pačioje trasoje metu.

Vairuotojo klasifikavimo ir vairavimo pobūdžio atpažinimo tyrimai [30] panaudojant inercinius CAN-bus transporto priemonės tinklo signalus. Tyrime parodyta didelė koreliacija tarp vairuotojo poveikio signalų (akseleratoriaus/stabdžių pedalo spaudimo, vairo kampo pasukimo) gautų iš CAN-bus, bei atskirai matuotų inercinių signalų (akselerometro bei giroskopo). Taip pat aprašyti tyrimai panaudojant signalus vairuotojui klasifikuoti.

Pasiūlytas dviejų etapų (klasterizavimo ir klasterizavimo) metodas vairavimo stiliui tyrime [29]. Tyrime naudojami priešais esančios transporto priemonės atstumo signalai išmatuoti lazerinio radaro bei inerciniai giroskopo ir akcelerometro signalai. Pirmiausia atliekamas transporto priemonės sekimo signalo klasterizavimas, o po to atliekamas klasterizavimas panaudojant inercinių signalų informaciją.

## 1.1. Išmaniųjų įrenginių kaip jutiklių platformos panaudojimas

Laikui einant, inercinių jutiklių kaip papildomų įrenginių atsirado nešiuojamuose įrenginiuose. Netrukus prasidėjo įvairių aplikacijų kūrimai, įgalintys panaudoti šių jutiklių informaciją. Neišimtis ir šių išmaniųjų įrenginių pritaikymas transporto sistemose. Dėl to jog visa sistemos platforma gali būti sukuriama tiesiog mobiliame įrenginyje, šių sistemų kaip platformų panaudojimas tapo pigiausias ir patogiausias: nereikalinga duomenų surinkimo ir kaupimo įranga, patogus GUI, duomenų persiuntimas ir integravimas su kitais įrenginiais. Pirmieji straipsniai panaudojantys išmaniųjų įrenginių inercinius jutiklius pasirodė 2008 metais.

Microsoft tyrimų centro mokslininkai [4] sukūrė sistemą pavadinimu „Nericell“, kuri panaudojama kelių būklei bei eismo spūsčių monitoringui panaudojant išmaniosius telefonus, panaudojant GPS, mikrofoną bei akcelerometrą. Kaip inercinis jutiklis panaudojamas trijų ašių akcelerometras. Straipsnyje aptariama kaip visa jutiklių informacija panaudojama atpažinti kelių nelygumams, duobėms, stabdymams, garsiniams signalams ir išmaniojo įrenginio lokalizavimui.

Sistema „Potholo Patrol“ [5] sukurta naudojant išmaniojo įrenginio platformą padeda nustatyti kelių paviršiaus kokybę. Kaip indikatorius panaudojamas trijų ašių akcelerometras, kurio informaciją pasitelkiant nustatomas kiek duobėtas kelias tam tikroje vietovėje, o GPS įrenginyje padeda tiksliai nustatyti šių kelio nelygumų lokaciją.

Apsvaigusio nuo alkoholio vairuotojo kelyje atpažinimui [7] buvo panaudotas išmaniojo įrenginio akcelerometras ir orientacinis jutiklis. Akcelerometras šiame tyrime panaudojamas stebėti kaip tiesiai vairuotojas važiuoja keliu. Dideli svyravimai informuoja apie nestabilią vairuotojo būseną ir pavojų eismo dalyviams. Sistema naudoja istorinius šablonus, kuriais remiantis lyginamas vairavimo pobūdis ir identifikuojamas apsvaigęs nuo alkoholio vairavimas.

Mobiliojo telefono taikymo ekologiškam vairavimui nustatyti [6] tyrimai buvo atliekami panaudojant GPS prietaisą ir akcelerometrą. Remiantis ekonomišką vairavimo standartais naudojant GPS ir akcelerometrą nustatinėjama ar standartai nėra pažeidžiami. Tam įvertinti buvo skaičiuojami įvairūs indeksai panaudojant sensorių informaciją.

Vairavimo stiliaus atpažinimo tyrimai naudojant išmanųjį telefoną [8] kada vairavimo stilius suskaidomas į tipinį (ne agresyvų) bei agresyvų buvo atlikti panaudojant išmanaus telefono magnetometrą, giroskopą, akcelerometrą, GPS bei kamerą. Buvo pritaikytas DTW (angl. Dynamic Time Warping) metodas jutiklių informacijai apjungti. Tyrimo metu sukurta sistema gebanti atpažinti tipinius ir agresyvius vairavimo manevrus panaudojant inercinių sensorių matavimus. Siekiant identifikuoti ar manevras agresyvus, jis lyginamas su duomenų bazėje sukauptais šablonais. Sukurtos sistemos pavadinimas – „MIROAD“ (angl. Mobile-Sensor-Platform for Intelligent Recognition Of Aggressive Driving).

Tyrime [9], panaudojant išmanųjį įrenginį bei transporto priemonės CAN-bus signalus efektyviam vairavimui kuro atžvilgiu nustatyti, buvo naudojami greičio, kuro sąnaudų, pagreičio, GPS ir kiti signalai. Išgauti signalai buvo panaudojami požymių išgavimui, kurie panaudojami klasifikavimui, o klasifikatoriaus signalas panaudojamas kaip priemonė suformuoti užuominą vairuotojui kaip keisti vairavimo pobūdį. Sukurta sistema apjungianti CAN-bus transporto priemonės jutiklių tinklą bei išmanųjį įrenginį buvo pavadinta „Driving Coach“.

Vairuotojo elgsenos nustatymui išmaniojo įrenginio panaudojimas buvo pasiūlytas tyrime [10], kuriame panaudojami akselerometro, magnetometro bei giroskopo signalai. Šiais signalais buvo nustatoma pozicijos, greičio, akseleravimo, stabdymo, pasisukimo kampo informacija, panaudojama tikimybiniam klasifikatoriui, kurio sprendimas nurodo kad vairuotojo elgesys saugus arba ne.

Keleivių komfortui transporto priemonėje nustatyti buvo sukurta sistema [11] pavadinimu „RESen“, kuri panaudoja išmanųjį telefoną. Tyrime buvo naudojami pagreičio ir gravitacijos jutikliai, o komfortabilaus važiavimo pojūtis suskaidomas į vertikalų ir horizontalų. Remiantis vairuotojo istorinėmis komfortabilaus vairavimo trajektorijomis sukurta sistema perduoda grįžtamąjį ryšį vairuotojui raskama nekomfortabilaus vairavimo anomalijas.

Išmanaus telefono trijų ašių akselerometro panaudojimas tirti saugų vairavimą buvo panaudotas tyrime [12]. Parodyta jog naudojant akselerometrą galima įvertinti automobilio būklę (pavarų perjungimas) ir kelio būklę (duobėtumą, paviršiaus nelygumą). Taip pat buvo nustatyta jog naudojant šį jutiklį galima identifikuoti staigius ir saugius vairuotojo maneversus (transporto priemonės akseleravimas, linijos kirtimas).

Siekiant palyginti išmanaus įtaiso jutiklių platformos galimybes su tradiciniu CAN-bus jutiklio tinklo informacijos galimybėmis transporto priemonėje automatiniam vairavimo manevrų atpažinimui buvo atliktas tyrimas [26]. Tyrimo rezultatai parodė jog įprasto CAN-bus tinklo jutiklių informacijos panaudojimas duoda prastesnius rezultatus lyginant su išmanaus telefono jutiklių platformos galimybėmis – panaudojans išmaniają planšetę pasiektas aukštesnis identifikavimo tikslumas. Tyrimo metu nesiekta palyginti jutiklių kokybę, o jutikliai buvo naudojami skirtingi, todėl rezultatai atspindi tik pačios platformos galimybes, o ne konkrečių jutiklių kokybę.

Vairuotojo elgesio profilio sudarymo naudojant išmanųjį telefoną tyrimas [13] buvo padarytas naudojant fuzzy išvadų darymo sistemą, kuri įvertina vairuotojo elgseną, kombinuojant skirtingus jutiklių fuzifikuotus duomenis. Tyrime buvo panaudoti pagreičio, orientacijos jutikliai bei GPS. Sistema remiantis turimais duomenimis nustato normalus, vidutinis ar agresyvus vairuotojo profilis.

Sistema, kuria galima nustatyti išmanaus įrenginio koordinačių sistemos ašių nuokrypio kampą nuo transporto priemonės koordinačių sistemos, sukurta tyrime [14]. Ši sistema kitaip nei dauguma sistemų naudojančių elektroninį kompasą – magnetometrą, naudoja IMU (inercinių matavimų jutiklius) ir GPS. Tyrime parodyta jog sistema gali apskaičiuoti kampą kai transporto priemonė yra bet kokioje pozicijoje įskainant ir kalnuotas vietoves.

Vairuotojo elgesiui įvertinti panaudojus mobilaus išmaniojo įrenginio jutiklius buvo atlikti tyrimai [15] ir sukurta sistema pavadinimu „MobiDriveScore“, kuri leistų vairuotojui pačiam įsivertinti savo varavimo pobūdį. Sistemoje panaudojami akselerometro bei GPS signalai. Remiantis šiais signalais vairuotojui įvertinti sudaromi rizikos indeksai, leidžiantys nuspręsti apie vairuotojo elgesį kelyje.

Viešojo transporto traukiniuose stovinčių keleivių diskomfortui matuoti buvo atlikti tyrimai [16] panaudojant išmanaus telefono akselerometro signalus. Darbe pasiūlyta diskomforto indekso apskaičiavimo metodika traukinio maršruto linijoms palyginti ir kokybei įvertinti.

Tyrimai susiję su vairavimo studijomis buvo atlikti [17] panaudojant išmaniojo telefono GPS, trijų ašių magnetometro ir akselerometro signalus. Tyrimo metu buvo palyginami išmanaus telefono jutiklių signalai su rinkoje egzistuojančių duomenų surinkimo prietaisų signalais. Buvo parodyta jog išmanaus telefono jutiklių signalų kokybė tolygi specialių tam skirtų įrenginių. Tai rodo jog nereikalinga speciali brangi įranga, o užtenka išmanaus telefono jutiklių vairavimo signalams tirti.

Mobilaus išmaniojo įrenginio jutikliai (GPS, giroskopo, akselerometro, magnetometro) hibridinėje elektrinėje transporto priemonėje panaudojami tyrime [27] vairuotojo elgesio monitoringo sistemai kurti.

## **2. AGRESYVIOS IR SAUGIOS VAIRUOTOJO ELGSENOS KLASIFIKAVIMAS**

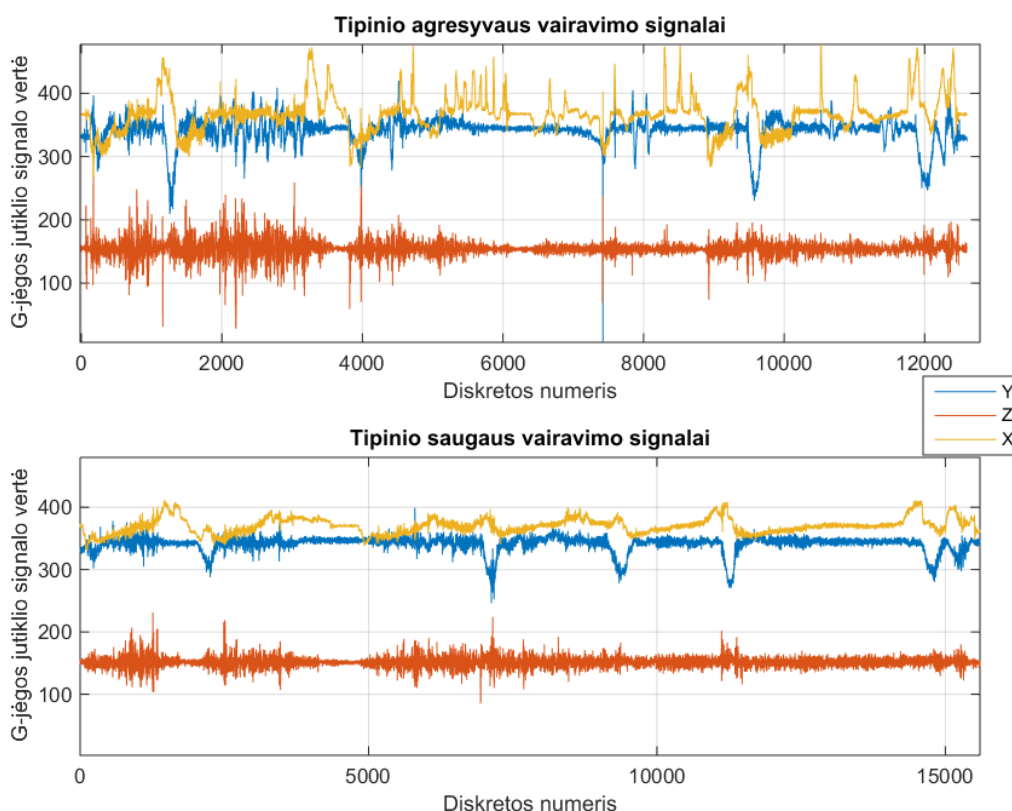
Skyriuje bus analizuojamas agresyvios ir saugios vairuotojo elgsenos klasifikavimas naudojant tam tikrus požymius kurie bus išgaunami iš trijų ašių G – jėgos jutiklio signalų. Jutiklio ašys atitinka išilgą (X), šoninę (Y) ir vertikalią (Z) pagreičio signalus. Tyrime naudojami viso 10 vairavimo duomenų takelių (trijų ašių G – jutiklio signalų), kai 6 iš jų yra agresyvaus vairavimo, o 4 – saugaus. Duomenys surinkti vairuojant lengvąjį automobilį. Važiavimai buvo atliekami tame pačiame maršrute važiuojant pastoviu (agresyviu ar saugiu) vairavimo stiliumi. Klasifikavimas atliekamas išgaunant požymius iš kiekvieno takelio duomenų intervalo, kai intervalo imties dydis lygus 1000 diskretų (diskretizavimo dažnis yra 17 Hz).

Toliau šis skyrius išskirtas į sekančias dalis: 1) Signalų pirminis apdorojimas; 2) Požymių išgavimas laiko ir dažnių srityse; 3) Agresyvaus ir saugaus vairavimo klasifikavimo metodai; 4) Požymių atrinkimo ir transformavimo metodai; 5) Klasifikavimo rezultatai.

## 2.1. Signalų pirminis apdorojimas

Signalų pirminis apdorojimas reikalingas siekiant sumažinti arba panaikinti pašalinės informacijos (triukšmų) įtaką iš signalo išgaunamai informacijai, bet tuo pačiu metu išlaikyti reikiamą informaciją tolimesnei informacijos išgavimo stadijai.

Tipiniai agresyvaus ir saugaus vairavimo trijų ašių G – jėgos jutiklio signalų pavyzdžiai pateikti 2.1.1 pav. Kaip matyti iš paveikslėlio, agresyviai vairavimui yra būdingos aukštesnio dažnio bei didesnių amplitudžių dedamosios signaluose. Taip pat matyti, jog abiem signalams būdingas tam tikras triukšmo lygis bei išskirtys (angl. outliers). Visi šiame darbe naudojami signalai diskretizuoti 17 Hz dažniu.



2.1.1 pav. Tipiniai agresyvaus ir saugaus vairavimo signalai

Siekiant įdėmiau išstudijuoti signalų harmoninių dedamųjų spektrą reikalinga atlikti spektrinę analizę. Tam pasiteklamos galios spektro tankio (angl. Power Spectral Density (PSD)) periodogramos. Periodograma yra neparimetrinė signalo galios spektro tankio įvertinimui skirta

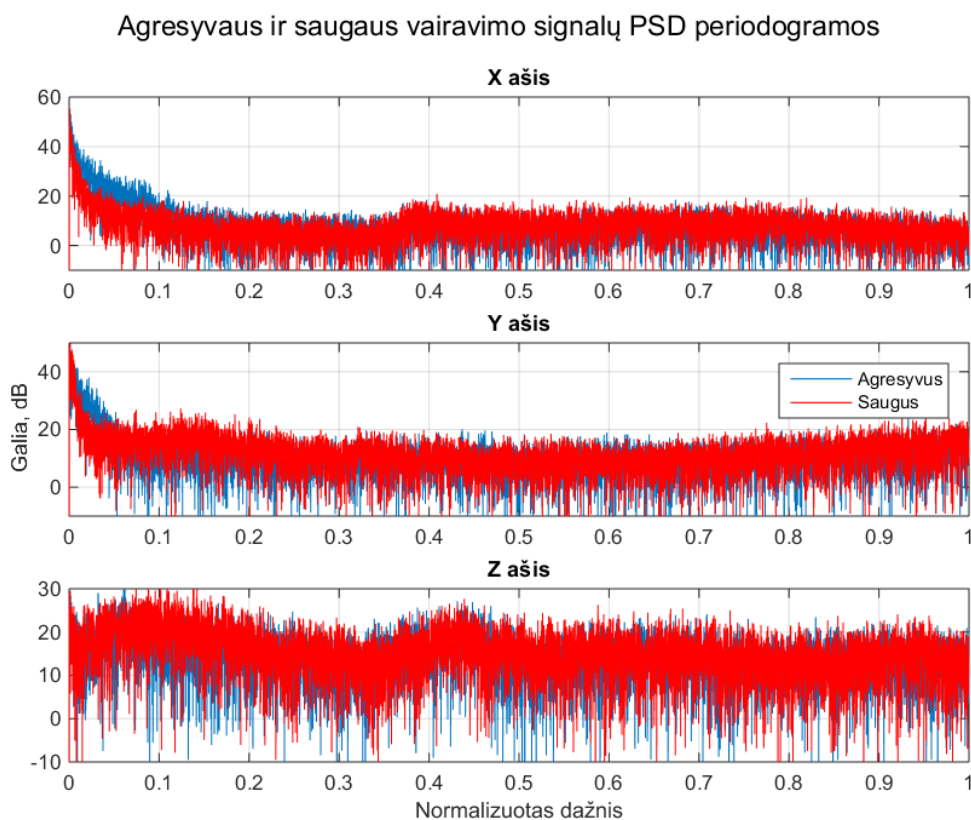
charakteristika. Tai yra signalo Furjė transformacija. Turint signalą  $x_n$  kurio diskretizavimo dažnis  $f_s$  per laiko vieneta, periodograma apibrėžiama kaip:

$$\hat{P}(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n e^{-i2\pi f n} \right|^2, \text{ kai } -\frac{1}{2\Delta t} < f \leq \frac{1}{2\Delta t}, \quad (1)$$

kur  $\Delta t$  – diskretizavimo laiko intervalas.

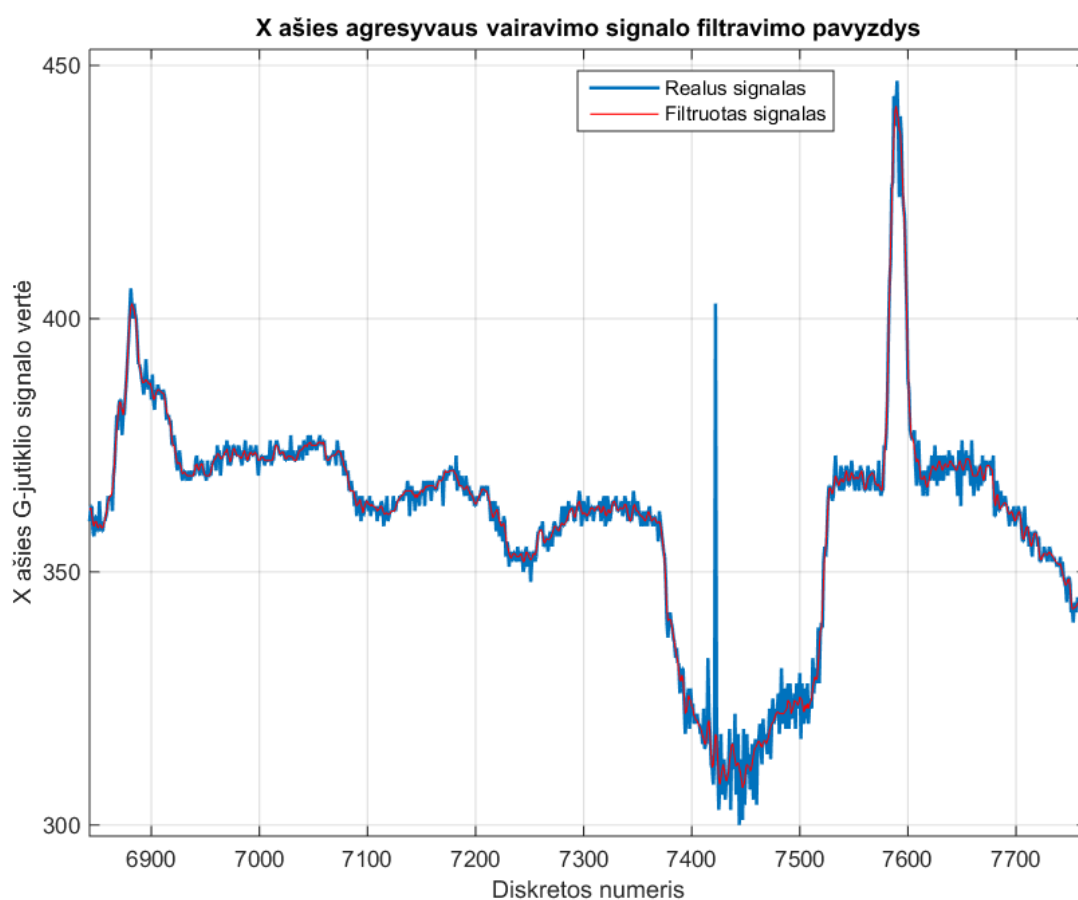
Kadangi vairavimo signalas fundamentaliai nėra periodinis, jo spektras nėra diskretinis, o juostinis. Ramaus ir agresyvaus vairavimo (2.1.1 pav.) signalų PSD periodogramos pavyzdžiai pateikti 2.1.2 paveiksle. Kaip matyti iš X ir Y ašies periodogramų, agresyvaus vairavimo žemų dažnių juosta yra platesnė, nei saugaus vairavimo. Tai rodo, jog agresyvaus X ir Y ašių signalų spektrai turi didesnę kiekį aukšto dažnio dedamųjų. Z ašies atveju žemų dažnių juosta labai neryški ir abiem atvejais panaši.

**PASTABA.** Prieš atliekant bet kokio signalo spektrinę analizę atliekamas signalo poslinkio ar trendo išminusavimas iš signalo. Tam atliekama signalo aproksimacija tiese naudojant mažiausių kvadratų metodą.



2.1.2 pav. Agresyvaus ir saugaus vairavimo signalų periodogramos kiekvienai signalo ašiai

Kaip matyti iš periodogramų, yra ganėtinai sudėtinga nustatyti optimalią dažnių ribą iki kurios informacija yra naudinga, o po kurios viską galima filtruoti taikant dažninį (pvz. FIR) filtrą, nes spektro juosta plati ir naudingo signalo informacija gali būti pasiskirsčiusi plačiame spektrų diapazone. Kitas, o taip pat ir paprastesnis būdas taikyti nedažninį filtravimą kombinuojant medianos ir vidurkio filtras: atliekamas trijų taškų slankios medianos filtravimas siekiant panaikinti siauras (vienos diskretos pločio) signalo išskirtis, o po to pritaikyti trijų taškų slankaus vidurkio filtrą, siekiant suglodonti filtruotą signalą. Filtrai pasirinkti trijų taškų, nes tai minimalus taškų kiekis mažiausiai iškraipantis originalų signalą, o taip pat pakankamas kiekis vienos diskretos pločio išskirtim pašalinti ir tuo pačiu sumažinamas triukšmo lygis signalė. Filtravimo pavyzdys pateiktas 2.1.3 paveiksle.



2.1.3 pav. Signalų filtravimas kombinuojant trijų taškų slankios medianos ir vidurkio filtras

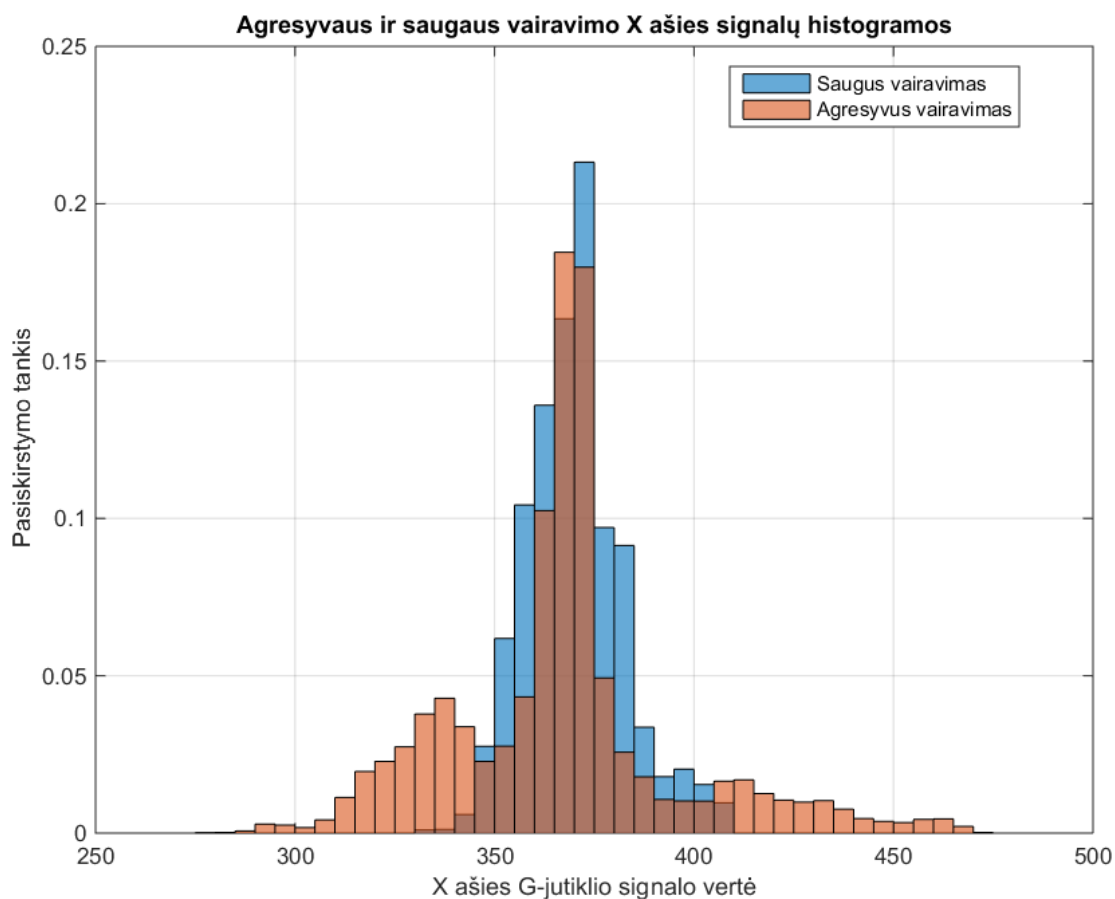


## 2.2. Požymių išgavimas laiko ir dažnių srityse

Šiame poskyriuje aptariami požymių išgavimo iš trijų ašių akselerometro signalų metodai. Visus šiuos metodus galima išskaidyti į dvi grupes: dažnių srities požymiai, bei laiko srities požymiai.

Toliau apibūdinami požymiai išgaunami laiko srityje iš kiekvieno filtruoto trijų ašių G – jėgos jutiklio signalo:

*Kvantiliai* (angl. quantiles) ir *atstumai tarp jų*. Skaičiuojamos signalo pasiskirstymo kvantilių vertės kas 0,1: nuo 0,1 iki 0,9. Taip pat skaičiuojami kvantilių atstumai kas 0,1 iš abiejų skirstinio pusių (tarp 0,2 ir 0,8 kvantilių ir t.t.). 2.2.1 paveiksle pateiktos agresyvaus ir saugaus vairavimo G – jutiklio X ašies signalų histogramos. Kaip matyti iš paveikslo, agresyvaus vairavimo pasiskirstymui būdinga didesnė variacija, o taipogi ir didesnės kraštinių kvantilių vertės.



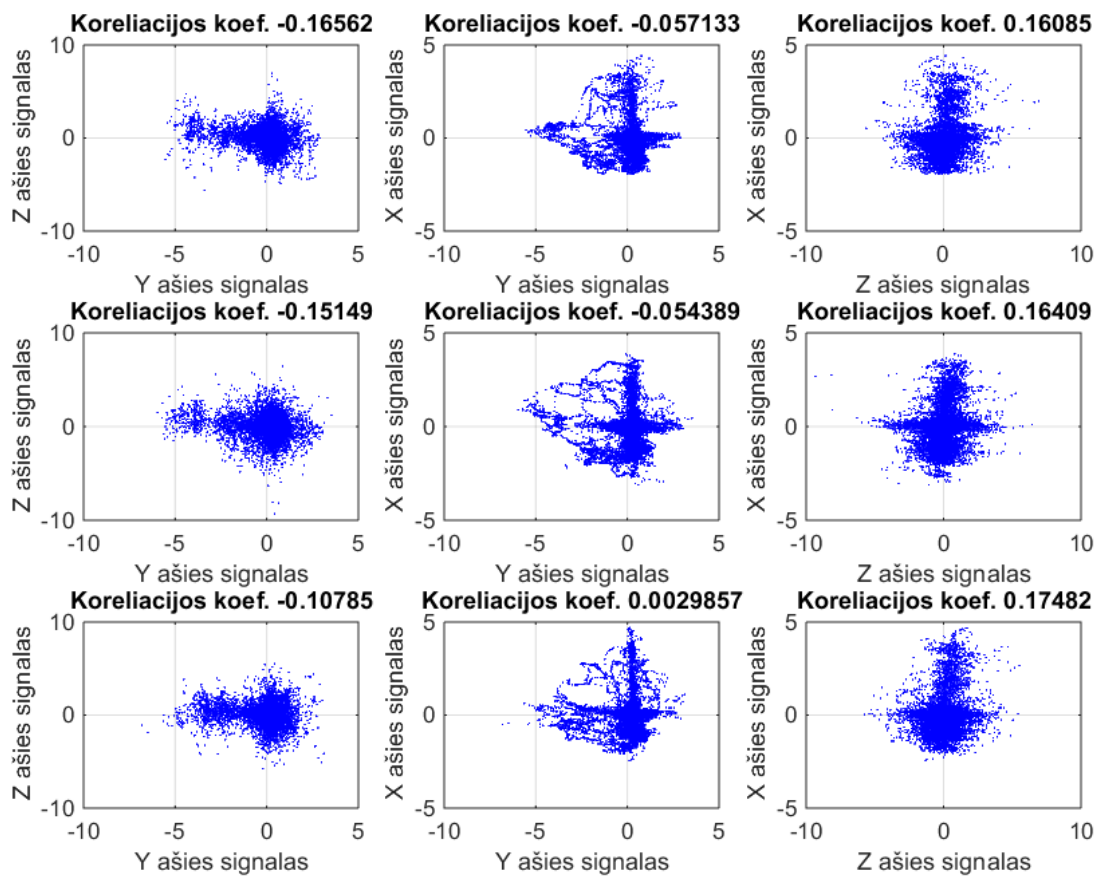
2.2.1 pav. Agresyvaus ir saugaus vairavimo G – jutiklio X ašies signalų histogramos

*Koreliacijos koeficientas tarp jutiklio signalų* (angl. correlation coefficient) nusako tiesinę priklausomybę tarp dviejų duomenų masių, apskaičiuojamas pagal formulę:

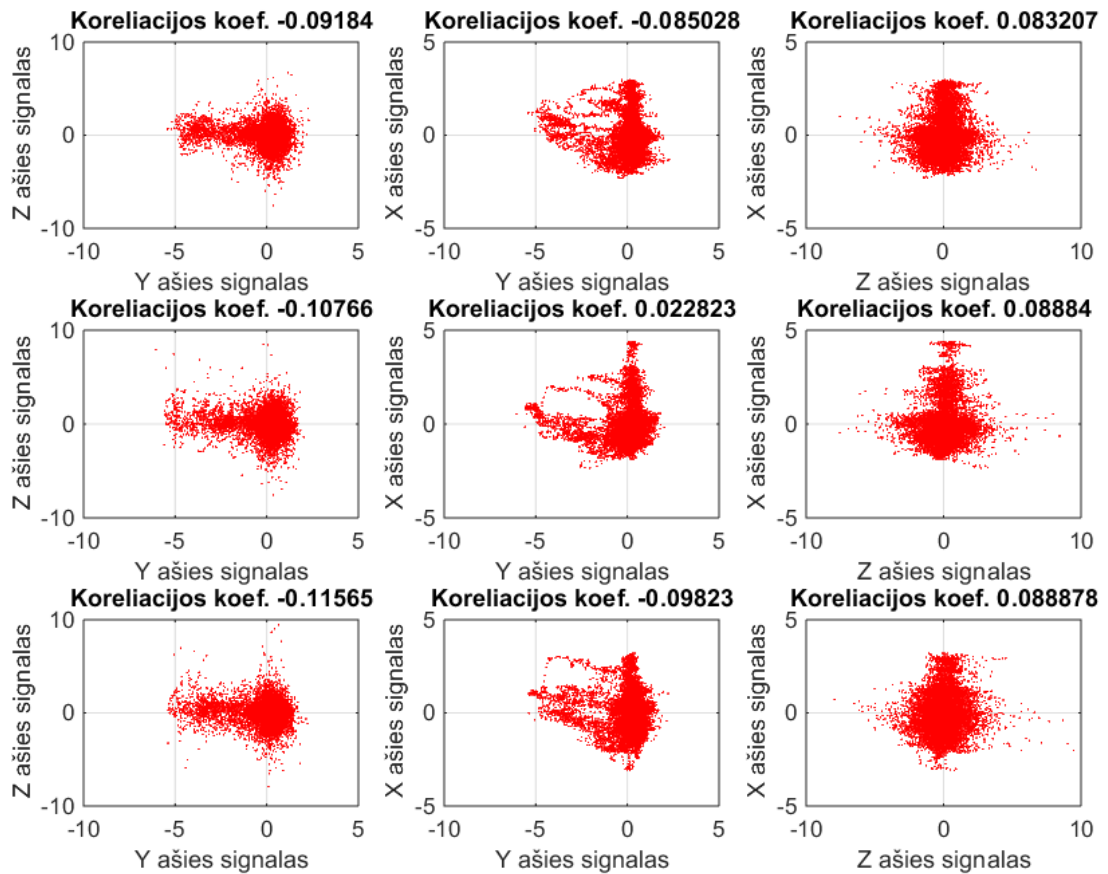
$$r_{xy} = \frac{1}{[(n-1) \cdot s_x \cdot s_y] \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}, \quad (2)$$

čia  $s_x, s_y$  – X ir Y važiavimų jutiklio ašių (X, Y ar Z) duomenų masyvų standartiniai nuokrypiai.

2.2.2 ir 2.2.3 paveiksluose pateikti atitinkamai agresyvaus ir saugaus vairavimų (po tris atskirus pilno maršruto važiavimo atvejus) signalų tarpusavio koreliacijos grafikai (grafikuose signalų vertės standartizuotos). Kaip matyti agresyvaus ir saugaus vairavimo atvejais koreliacijos koeficientai skiriasi.

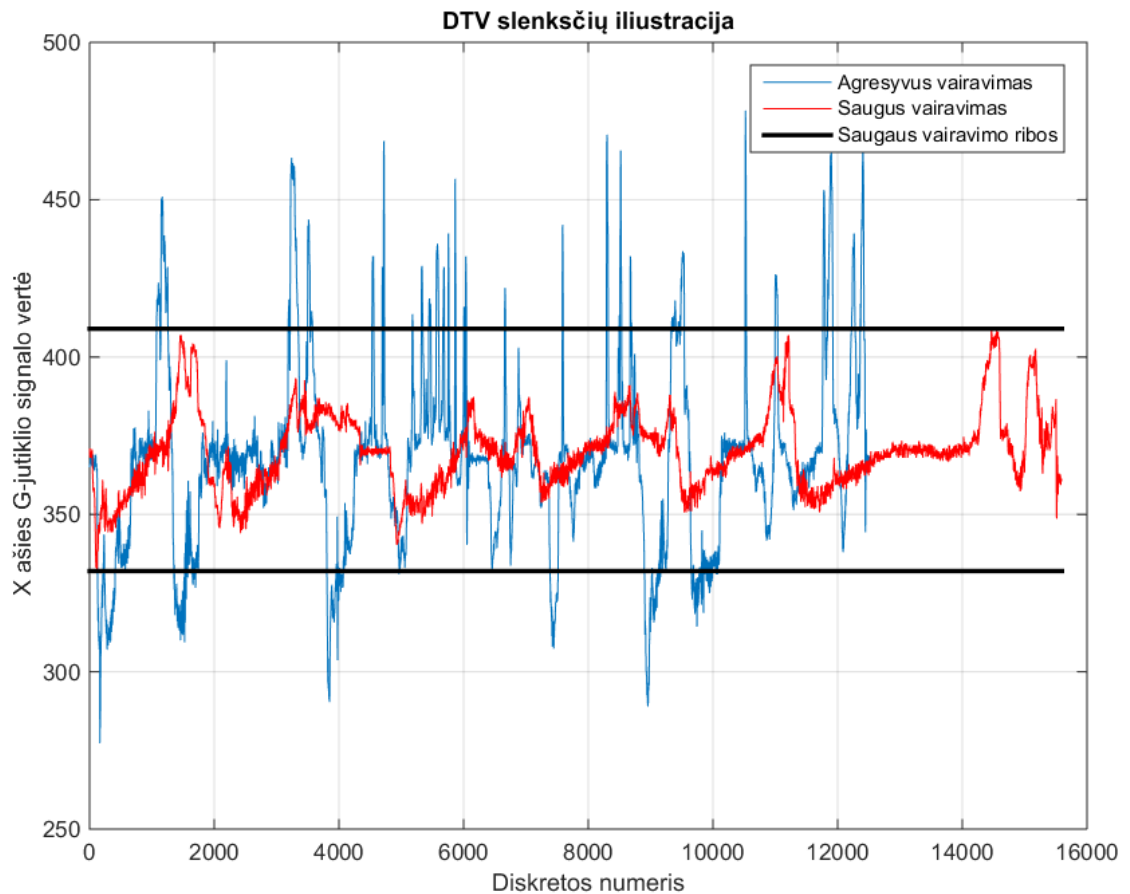


2.2.2 pav. Trijų agresyvaus vairavimo pilnų važiavimų atvejų signalų tarpusavio koreliacijų pavyzdžiai



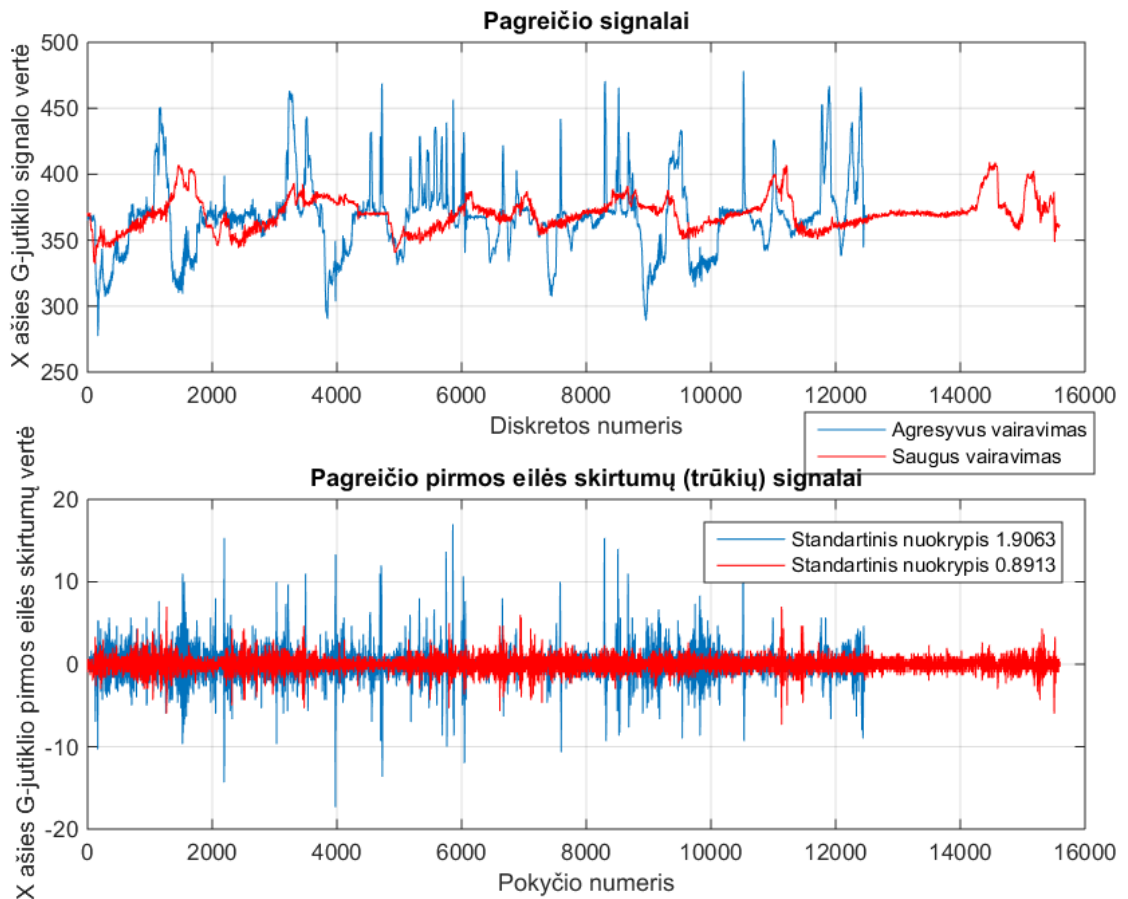
2.2.3 pav. Trijų saugaus vairavimo pilnų važiavimų atvejų signalų tarpusavio koreliacijų pavyzdžiai

*Duomenų ribos viršijimo intensyvumo skaičiavimas* (angl. Data Threshold Violation (DTV)). Skaičiuojamas užsiduotų signalo ribų viršijimo procentas, kuris padauginamas iš viršijimo amplitudinės vertės: dažnas, bet palyginti žemų amplitudžių viršijimas rodo mažensį agresyvumą nei dažnas, bet palygint aukštų amplitudžių viršijimas. Slenksčių nustatymo pavyzdys naudojant saugaus vairavimo signalą kaip etaloną pateiktas 2.2.4 paveiksle.



2.2.4 pav. G – jutiklio X ašies signalo DTV slenksčių nustatymo pavyzdys

*Signalų pirmos eilės skirtumų (trūkių) standartinio nuokrypio skaičiavimas (angl. jerk).*  
 Fizikoje trūkis apibrėžiamas kaip pagreičio kitimo greitis. Šiame tyrime trūkiai skaičiuojami kaip pirmos eilės filtruoto signalo skirtumai. Trūkiai yra tiesiogiai susiję su vairuotojo agresyvumu, kuro sąnaudomis ar keleivio komfortu. 2.2.5 paveiksle pateikti trūkių palyginimai vairuojant agresyviai ir saugiai.



2.2.5 pav. Agresyvaus ir saugaus vairavimo pagreičio pirmos eilės skirtumų (trūkių) palyginimas

Dažnių srityje bus išgaunami požymiai iš spektrinės charakteristikos. Prieš tai jau buvo užsiminta jog vairavimo agresyvumas atsispindi signalo spektre (2.1.2 pav.). Siekiant išsamiau analizuoti spektro signalo įtaką tam tikro stiliaus vairavimo laikotarpiu toliau pasitelkiama spektrograma. Spektrograma nusako signalo dažnines savybes tam tikrame laiko intervale. Ji gaunama atliekant svertinę signalo Furjė transformaciją tam tikrame laiko intervale:

$$F(\omega, \tau) = \int f(t)w(t - \tau)e^{-j\omega t} dt , \quad (3)$$

čia  $\omega = 2\pi f$  – dažnio įvertis;

$\tau$  – laiko intervalas;

$w(t - \tau)$  – transformacijos svorinė (lango) funkcija (šiam darbe naudojama Hammingo lango funkcija).

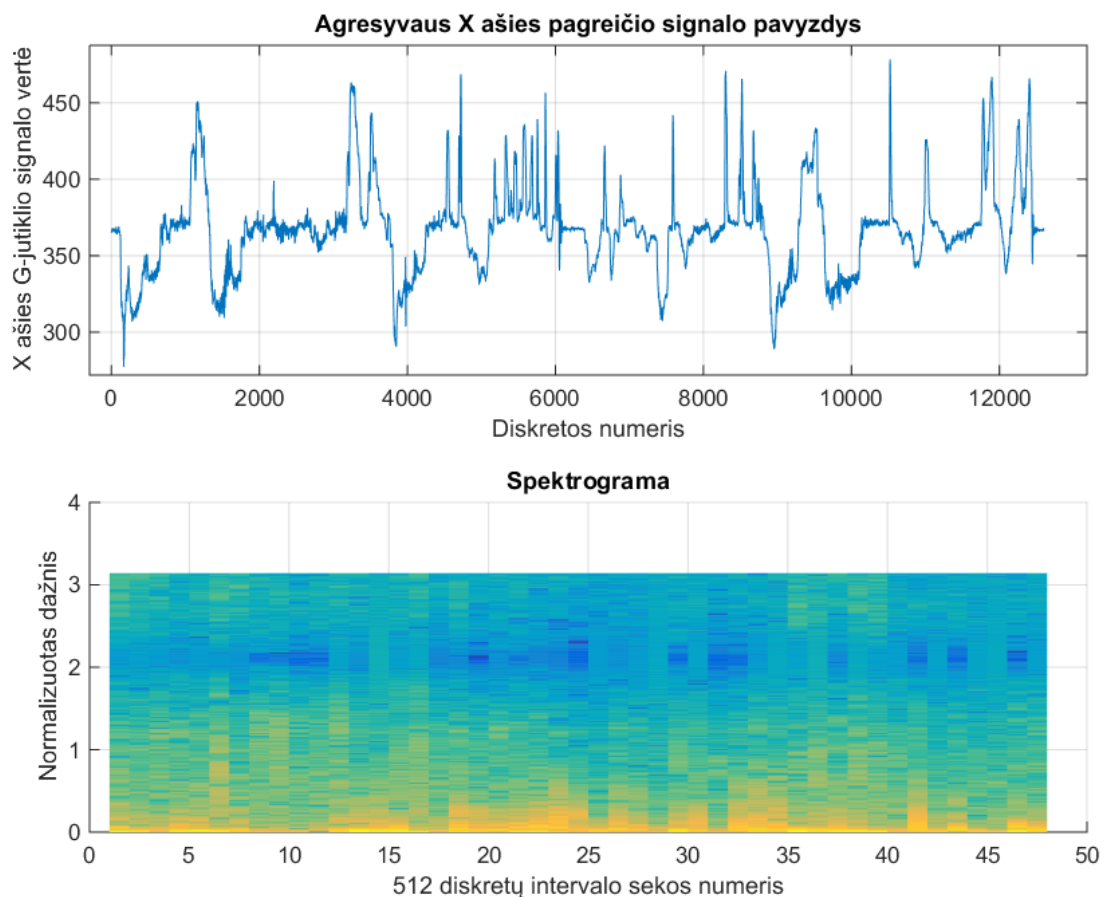
Po svertinės signalo Furjė transformacijos kiekvienam intervalui skaičiuojamas vienpusis galios spektro tankis:

$$P(\omega, \tau) = k|F(\omega, \tau)|^2, \text{ kai } k = \frac{2}{\sum_{n=1}^L 2 \cdot |w(n)|} \quad (4)$$

čia  $L$  – lango ilgis;

$F(\omega, \tau)$  – Furjė transformacija.

Šiuo atveju spektrograma leidžia palyginti signalo dažnines savybes laike ir geriau suprasti agresyvaus vairavimo įtaką signalo spektrui tam tikru periodu. 2.2.6 paveiksle pateiktas agresyvaus vairavimo X ašies G – jutiklio signalo spektrogramos pavyzdys. Spektrogramos spalva rodo tam tikro dažnio galios intensyvumo logaritmo dydį (geltonas – didelis, mėlynas – mažas). Kaip matyti iš spektrogramos, esant agresyviai vairavimui (pvz., signalo intervalas ~5000 – 6000 diskretos), spektro galia ties aukštesniais dažniais santykinai reikšmingai išauga.



2.2.6 pav. Filtruoto agresyvaus signalo spektrogramos pavyzdys

Šiame darbe kaip dažnių srities požymis bus naudojamas santykinis galios spektro periodogramos intensyvumas tam tikroje dažnių juostoje. Visas galios spektras suskaidomas į 10 spektrinių juostų, kiekvienos juostos galios intensyvumai susumuojami ir padalinami iš visos

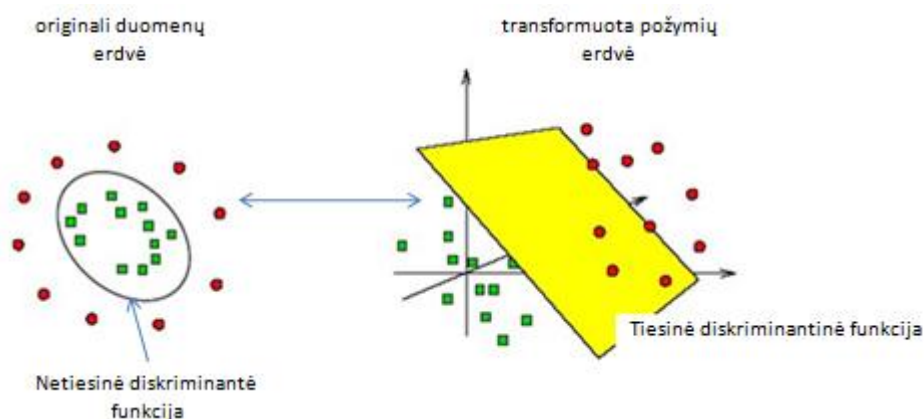
spektro galios intensyvumų sumos. Taip gaunama santykinė spektro galia tam tikroje dažnių juostoje.

### 2.3. Agresyvaus ir saugaus vairavimo klasifikavimo metodai

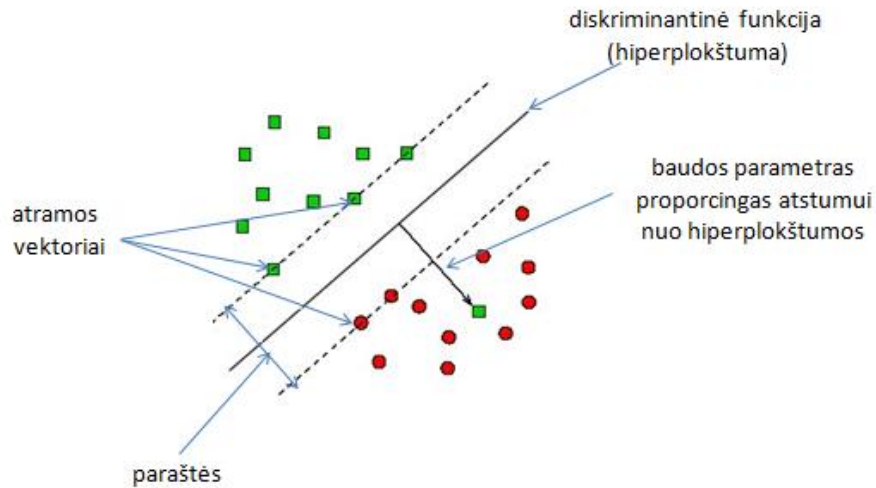
Šio tyrimo metu bus naudojama keletas klasifikavimo algoritmų siekiant atlikti klasifikavimo tikslumo įvertinimą bendruoju atveju, nepriklausomai nuo klasifikavimo algoritmo. Tyrime bus naudojami atsitiktinių miškų (angl. Random Forest (RF)), neuroninių tinklų (angl. Neural Networks (NN)), atraminių vektorių mašinos (angl. Support Vector Machines (SVM)), bei  $k$  – artimiausių kaimynų (angl.  $k$  – Nearest Neighbours ( $k$ NN)).

#### 2.3.1. Atraminių vektorių mašina

Atraminių vektorių mašinos (angl. Support Vector Machine) tai tiesinis klasifikavimo algoritmas kuris tiesiškai neišskiriamas klases įėjimo požymių erdvėje tiesiškai atskiria transformuotoje hipererdvėje (2.3.1.1 pav.) panaudojant branduolio triuką (angl. kernel trick) pasirenkant tam tikrą netiesinę branduolio funkciją (angl. kernel function). Todėl požymių erdvėje diskriminantinė funkcija gali įgauti netiesinę formą. Šis klasifikavimo algoritmas originaliai pritaikytas išskirti dvi klases. Kai duomenyse didesnis klasių skaičius naudojama viena prieš vieną arba viena prieš visas klasių atskirimo metodika. Šiuo atveju algoritmas idealiai tinkamas, nes nagrinėjamas dviejų klasių (agresyvaus ir saugaus vairavimo) klasifikavimo uždavinys. Klasifikavimo algoritmo tikslas – išskirti klases su didžiausiomis paraštėmis (angl. margins) kaip pateikta 2.3.1.2 pav. Kada klasės tiesiškai neatskiriamos įvedamas baudos parametras proporcingas atstumui nuo diskriminantinės funkcijos.



2.3.1.1 pav. Atraminių vektorių netiesinės diskriminantinės funkcijos gavimo pavyzdys



2.3.1.2 pav. Atraminų vektorių klasifikavimo metodas tiesiškai neišskiriamoms klasėms įvedant baudos parametą

Tiesiškai neatskiriamų klasių problema pirminė optimizavimo problema:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i ;$$

$$\text{kai } \begin{cases} y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ \xi_i \geq 0, \quad i = 1, \dots, n. \end{cases} \quad (5)$$

Čia  $w$  – parametų vektorius;

$b$  – poslinkio parametras;

$C$  – išorinis baudos parametras (angl. cost parameter);

$\xi$  – duomenų taško atstumo nuo hiperplokštumos kintamasis (angl. slack variable), skirtas persidengiančių klasių uždaviniui spręsti;  $n$

$n$  – duomenų stebėjimų skaičius;

$\phi$  – branduolio funkcija (tyrime bus naudojama Gauso branduolio funkcija);

$y$  – klasės (išėjimo) vertė (1 arba -1).

Antrinė optimizavimo problema formuluojama taip:

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha ;$$

$$\text{kai } \begin{cases} y^T \alpha = 0, \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n. \end{cases} \quad (6)$$

Čia  $e$  – vienetinis vektorius (visos vektoriaus vertės lygios 1);

$Q$  –  $n \times n$  branduolio matrica, kurios nariai  $Q_{ij} \equiv K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ ;



$\alpha$  – Lagranžo daugiklių vektorius (duomenų stebėjimų taškus atitinkančios vertės, kurių nenulinės vertės atitinka atraminių vektorių pozicijas originalių įėjimo požymių erdvėje). Atraminių vektorių modelio diskriminantinė funkcija gaunama:

$$\text{sgn}(\sum_{i=1}^n y_i \alpha_i K(x_i, x) + \rho). \quad (7)$$

Čia  $\text{sgn}(\cdot)$  – ženklų funkcija.

Antrinei atraminių vektorių mašinos optimizavimo problemai spręsti naudojamas SMO (angl. Sequential Minimal Optimization) algoritmas. Šis optimizavimo algoritmas yra iteracinis, SMO suskaido antrinę optimizavimo problemą į eilę mažesnių problemų, kurios sprendžiamos analitiškai. Dėl tiesinių apribojimų, kurie susiję su Lagranžo daugiklių naudojimu  $\alpha_i$ , mažiausia galima problema susideda iš dviejų šių daugiklių. Tada bet kuriems dviems Lagranžo daugikliams, apribojimai supaprastinami iki:

$$\begin{cases} 0 \leq \alpha_i, \\ \alpha_j \leq C, \\ y_i \alpha_i + y_j \alpha_j = k. \end{cases} \quad (8)$$

Čia  $k$  – neigiama visų kitų likusių narių suma (lygybės apribojime), kuri kiekvienos iteracijos metu yra pastovi.

Ši problema jau išsprendžiama analitiškai: reikalinga surasti vienos dimensijos kvadratinės funkcijos minimumą. Nors algoritmas konverguoja, tam kad paspartinti konvergavimo greitį Lagranžo daugiklių poros parenkamos naudojant tam tikras euristines taisykles.

Išorinių parametrų optimizavimas ( $C$  bei kitų parametrų priklausomai nuo pasirinktos branduolio funkcijos) atliekamas naudojant kryžminę validaciją.

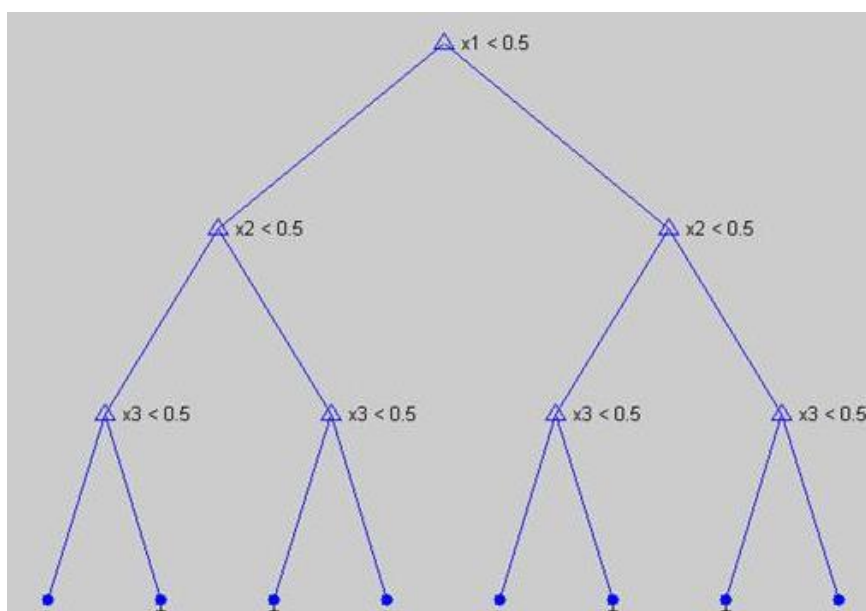
Dėl to jog požymiai transformuojami į daugiadimensę erdvę optimaliam skiraimajam hiperpaviršiui (hiperplokštumai) nustatyti, šis modelis pernelyg nėra jautrus duomenų daugiadimensiškumui.

### 2.3.2. Atsitiktinių miškų klasifikavimo algoritmas

Šis algoritmas yra sudarytas iš daugybės paprastų klasifikatorių – klasifikavimo medžių, todėl dar vadinamas klasifikavimo medžių (angl. decision tree) kolektyvu. Atsitiktinį mišką (angl. Random Forest (RF)) sudaro daug medžių užaugintų ant atsitiktinai parinktų požymių,

taip pat medis apmokomas ne su visais atsitiktinai parinktų požymių duomenimis, bet taip pat su atsitiktinai parinktais duomenimis. Taikant tokią metodiką ir naudojant daug užaugintų medžių vidutinis tikslumas lyginti su vienu sprendimo medžiu pagerėja. Dėl minėtų atsitiktinių duomenų taškų bei požymių parinkimų ir daugybės medžių šio klasifikatoriaus sprendimai linkę mažiau persimokyti ir duoti statistiškai patikimesnius rezultatus kuo medžių skaičius didesnis. Klasifikavimo medžio pavyzdys pateiktas 2.3.2.1 pav. Kiekvienas mazgas atitinka tam tikro kintamojo (požymio) stebėjimo reikšmės tikrinimą pagal sąlygą, o paskutiniai mazgai – lapai, tai klasės priskyrimas duomenų taškui po visų sąlygų nuo viršaus iki apačios tikrinimų.

Taip pat naudojant šį algoritmą papildomas kryžminis modelio validavimas modelio stabilumui patikrinti nėra būtinas dėl minėtų savybių, o tikslumas patikrinamas naudojant duomenis nepaimtus mokymo metu kaip testavimo ir nustatant klasifikavimo klaidą (angl. Out-of-Bag Error (OOBError)). Dėl savo statistinių savybių modelis gali gerai patikimai naudojant visą požymių erdvę.



2.3.2.1 pav. Klasifikavimo į dvi klases medžio pavyzdys

### 2.3.3. Neuroninių tinklų klasifikatorius

Dirbtiniai neuroniniai tinklai plačiai naudojami tiek klasifikavimo tiek regresijos uždaviniams spręsti. Tiesioginio sklidimo dirbtinio neuroninio tinklo (angl. Feed-Forward Neural Network (FFNN)) architektūra detalizuota 2.3.3.1 pav. Tai netiesinis klasifikavimo algoritmas. Dirbtinis neuroninis tinklas susideda iš tam tikro sluoksnių skaičiaus: įėjimo sluoksnio,  $k$  paslėptųjų sluoksnių bei išėjimo sluoksnio. Kiekvienas įėjimas  $x_1, x_2, \dots, x_r$ , atitinka

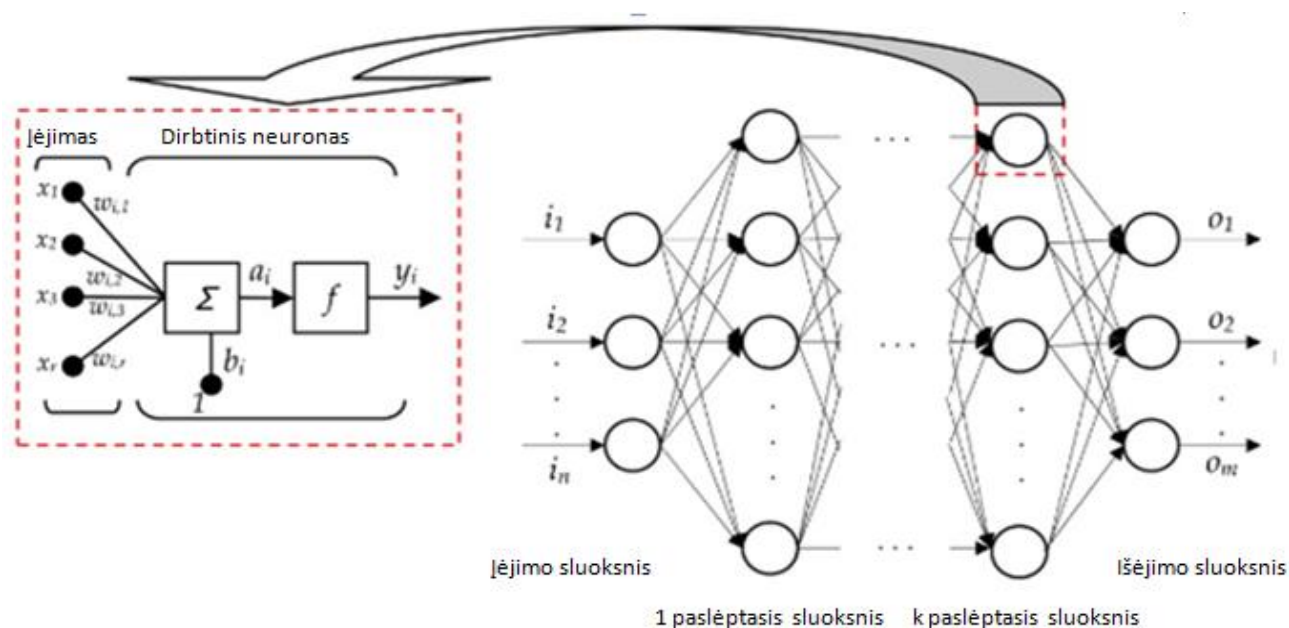
tam tikrą požymį su tam tikru stebėjimų skaičiumi, kurie pasveriami atitinkamai svoriais  $w_{i,1}$ ,  $w_{i,2}, \dots w_{i,r}$  kurių vertės nustatomos mokymo metu ir šių pasvertų stebėjimų suma yra tam tikros netiesinės aktyvavimo funkcijos įėjimas. O tolesnio sluoksnio įėjimai yra prieš tai buvusio sluoksnio išėjimai  $y_i$  ir taip toliau. Klasifikavimo uždaviniui spręsti išėjimo sluoksnio mazgų skaičius turi sutapti su klasių skaičiumi. Stebėjimas priskiriamas tai klasei kuriai atitinkančio išėjimo sluoksnio mazgo išėjimas  $o_i$  duoda didžiausią aktyvavimo reikšmę (aktyvavimo funkcijos išėjimą).

Apmokymo metu sureguliuojami neuroninio tinklo svoriai taip kad būtų minimizuojamas tam tikras baudos matas (pvz. suminė kvadratinė klaida, ar klasifikavimo klaida) naudojant gradientinio optimizavimo (atgalinio klaidos sklidimo (angl. backpropagation)) algoritmus.

Pagrindinis neuroninio tinklo trūkumas tai kad reikalingas palyginti su kitais klasifikavimo algoritmais didelis duomenų kiekis siekiant gauti geras tinklo apibendrinimo savybes, bei dėl to jog naudojama atsitiktinė tinklo svorių inicializacija galimas algoritmo persimokymas, todėl reikalingas rezultatų vidurkinimas ar kitoks statistinis įvertinimas naudojant keletą neuroninių tinklų vienu metu. Neurono išėjimas aprašomas tokia lygtimi:

$$y_i = f(w_{i,1}x_1 + w_{i,2}x_2 + \dots + w_{i,r}x_r + b) \quad (9)$$

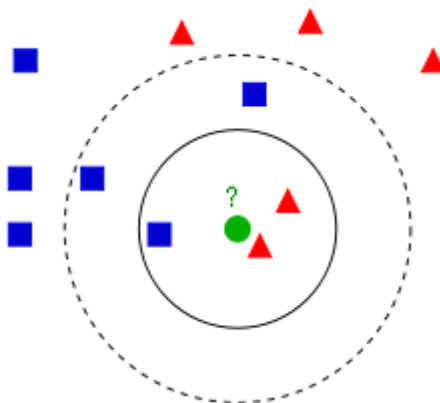
Čia  $f(\cdot)$  – tam tikra neuroninio tinklo aktyvavimo funkcija (sigmoidinė, radialinė bazinė, tiesinė, tangentinė ir kt.).



2.3.3.1 pav. Tiesioginio sklidimo neuroninio tinklo architektūra

### 2.3.4. $k$ – artimiausių kaimynų klasifikatorius

$k$  – artimiausių kaimynų (angl.  $k$  – Nearest Neighbors ( $k$ NN)) klasifikatorius priskirdamas naują duomenų tašką tam tikrai klasei remiasi  $k$  artimiausių (artumo atstumas nustatomas pagal pasirinktą metriką) kaimyninių taškų informacija (jų klasėmis) ir priskiria naują duomenų tašką tai klasei kurios klasės balsų skaičius (kaimyninių taškų skaičius iš visų  $k$  artimiausių kaimynų) yra didžiausias. Grafikas iliustruojantis šio klasifikatoriaus darbą pateiktas 2.3.4.1 pav. kai žalias duomenų taškas priskiriamas pagal  $k$  artimiausių kaimynų klasių balsus.



2.3.4.1 pav.  $k$  – artimiausių kaimynų klasifikavimo algoritmo veikimas

## 2.4. Požymių atrinkimo ir transformavimo metodai

Kitas svarbus uždavinys susijęs su agresyvaus ir saugaus vairavimo stiliaus klasifikavimu uždavinys yra požymių atrinkimas. Kai kurie klasifikavimo algoritmai (pvz.  $k$ NN) gali būti pernelyg jautrūs duomenų dimensiškumui (angl. curse of dimensionality), tai lemia prastus klasifikavimo rezultatus. Kita problema susijusi su per dideliu nereikalingų požymių kiekių naudojimu, kas turint palyginti mažai stebėjimo duomenų taškų gali klaidinti klasifikatorių ir taip pabloginti prognozavimo tikslumą. Nereikalingi požymiai gali būti dvejopi: požymis gali būti nereikalingas jei jis labai koreliuoja su kitais požymiais (koduoja tą pačią informaciją) arba jis neturi klasių skirtumo informacijos (tolygus triukšmui). Taip pat požymių atrinkimo nauda susijusi su mažesnių skaičiavimų kiekiu: nereikalinga naudoti tam tikrus požymius. Siekiant atrinkti ir palyginti požymių atrinkimo efektyvumą taikomi keletas požymių atrinkimo algoritmų, kurie nepriklausomai nuo klasifikatoriaus geba suranguoti požymius nuo informatyviausio iki mažiausiai informatyvaus.

Požymių transformavimas – tai kitas būdas požymių dimensiškumui bei kiekiui sumažinti. Šiuo atveju požymiai nėra atrenkami, bet jie yra transformuojami iš daugiadimensinės

erdvės į mažesnių dimensijų erdvę pagal tam tikrus kriterijus kurie gali pagerinti klasifikavimo uždavinio sprendimą.

**PASTABA:** siekiant suvienodinti požymių svorius prieš požymių atrinkimo uždavinį pirmiausia atliekamas kiekvieno iš požymių normalizavimas (per visas klases) pagal sekančią formulę:

$$z = \frac{x - \bar{x}}{s}. \quad (10)$$

Čia  $x$  – požymio duomenų taškas;

$\bar{x}$  – požymio duomenų vidurkio įvertis;

$s$  – požymio duomenų standartinio nuokrypio įvertis.

#### 2.4.1. Porinis Stjudento $t$ – testas

Taikant šį metodą įvertinama kiekvieno požymio svarba objektyviai naudojant porinį Stjudento  $t$  statistinį kriterijų. Tai parametrinis testas kuriuo metu palyginami dviejų nepriklausomų duomenų imčių (šiuo dviejų atveju klasių kiekvieno požymio stebėjimų) išsidėstymo parametrai – centrai. Yra iškeliami nulinė hipotezė kad du duomenų vektoriai priklauso atsitiktinių dydžių imtims iš to pačio normalinio pasiskirstymo dėsnio su lygiais vidurkiais ir lygiomis bet nežinomomis variacijomis, t.y. kad duomenys kilę iš to pačios populiacijos. Atlikus hipotezės tikrinimą yra gaunamos tikimybinės reikšmės (angl.  $p$  – values), kurio parodo hipotezės tikėtumą tarp duomenų vektorių kiekvienam požymiui, kitaip tariant, didelė  $p$  – vertė parodo kad labai tikėtina jog duomenys yra kilę iš to pačio skirstinio. Tai tuo pačiu parodo jog šių duomenų požymių tarpusavyje klasifikuoti neišeis (duomenys neišsiskirstę į dvi atskiriamas grupes). Kuo mažesnė požymio  $p$  – vertė tuo labiau tikėtina jog duomenų požymį bus galima klasifikuoti. Praktikoje plačiai naudojamas  $p=0,05$  reikšmingumo lygmuo: požymių reikšmės kurių  $p$  – vertės mažesnės už 0,05 priskirtinos kaip neatitinkančios hipotezės (tinkamos klasifikuoti), o didesnės – atvirkščiai. Jei hipotezė patvirtinama, tai  $t$  – pasiskirstymas yra tapatus Stjudento skirstiniui. Statistinis testas yra atliekamas su  $t$  įverčiais apibrėžiamais formule:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}, \quad (11)$$

čia  $\bar{x}, \bar{y}$  – požymio dviejų imčių (agresyvaus ir saugaus vairavimo klasių) vidurkiai;

$s_x, s_y$  – požymio dviejų imčių standartinio nuokrypio įverčiai;

$n, m$  – požymių dviejų imčių dydžiai.

### 2.4.2. Santykinė entropija

Santykinė entropija (angl. relative entropy) dar kitaip žinoma kaip Kullback – Leibler divergencija arba atstumas (angl. Kullback – Leibler divergence/distance) yra matas nusakantis entropinį atstumą tarp dviejų skirstinių. Informacijos teorijoje apibrėžiama kaip informacijos skirtumu kada vienas skirstinys yra panaudojamas vietoj kito. Santykinė entropija apibrėžiama tokia formule:

$$KL(X; Y) = \sum_k p_k(X) \log \left( \frac{p_k(X)}{p_k(Y)} \right), \quad (12)$$

kur  $p(X), p(Y)$  – diskretiniai duomenų skirstiniai.

Santykinės entropijos matas panaudojamas entropiniam atstumui tarp požymio klasių porų verčių skaičiuoti. Kuo didesnė entropija tuo požymiai kiekvienoje klasėje skiriasi labiau ir tikėtina kad bus labiau tinkami klasifikavimui. Jei entropija maža tai požymiai yra panašūs ir netinkami atskiroms klasėms išskirti.

Šis metodas turi panašius trūkumus kaip ir Stjudento  $t$  – testas (porinis), bet yra fundamentaliai skirtingas nuo pastarojo.

**PASTABA.** Šiame tyrime santykinė entropiją skaičiuojama kiek supaprastintai, nes priimama prielaida apie normalųjį duomenų pasiskirstymą. Tyrime naudojama santykinės entropijos funkcija kai skirstiniai yra normaliniai aprašoma pagal formulę:

$$KL_{ij} = \frac{1}{2} \left( \frac{\sigma_i^2}{\sigma_j^2} + \frac{\sigma_j^2}{\sigma_i^2} - 2 \right) + \frac{1}{2} (\mu_i - \mu_j)^2 \left( \frac{1}{\sigma_i^2} + \frac{1}{\sigma_j^2} \right), \quad (13)$$

kur  $\sigma$  – požymio duomenų standartinis nuokrypis;

$\mu$  – požymio duomenų vidurkis.

### 2.4.3. Bhattacharyya atstumas

Bhattacharyya atstumas matuoja skirtumą tarp duomenų tikimybės tankių porų. Šiuo metodu matuojamas tikimybių tankių persidengimas – minimali klasifikavimo klaida naudojant Bajeso klasifikatorių. Ši klaida teoriškai apibrėžiama formule:

$$P_e = \int_{-\infty}^{\infty} \min[P(\omega_i)p(x|\omega_i), P(\omega_j)p(x|\omega_j)]dx . \quad (14)$$

Kadangi analitinis šio integralo sprendinys bendru atveju nėra tiksliai išgaunamas, nustatoma viršutinė šio sprendinio riba vadinama Chernoff riba (angl. Chernoff bound), kuri, priimant prielaidą apie normaliuosius skirstinius, gaunama kaip:

$$\epsilon_{CB} = \sqrt{P(\omega_i)P(\omega_j)\exp(-B)}, \quad (15)$$

kur,

$$B = \frac{1}{2} \left( \frac{(\mu_i - \mu_j)^2}{(\sigma_i + \sigma_j)/4} + \log \left( \frac{\sigma_i + \sigma_j}{(\sqrt{\sigma_i \sigma_j})/2} \right) \right). \quad (16)$$

$B$  yra apibrėžiamas kaip Bhattacharyya atstumas. Šio metodo trūkumai ir privalumai iš esmės tie patys kaip ir santykinės entropijos ar Stjudento  $t$  – testo metodų, aptartų anksčiau.

#### 2.4.4. Wilcoxon rangų sumos testas

Wilcoxon rangų sumos (angl. Wilcoxon rank – sum) arba Mann – Whitney  $U$  – testas yra neparametrinis statistinės hipotezės tikrinimo testas naudojamas palyginti dviejų imčių duomenų populiacijas. Jo taikymas yra kaip alternatyva poriniam Stjudento  $t$  – testui, kada negalima priimti prielaidos apie normalųjį duomenų pasiskirstymą. Šio testo metu tikrinamas duomenų medianų skirtumo reikšmingumas (Stjudento  $t$  – testo atveju lyginami vidurkiai), t.y., iškeliamą hipotezė jog dviejų duomenų imčių medianos yra lygios. Kaip ir Stjudento  $t$  – testo atveju požymiams indeksuoti naudojamos statistinės hipotezės reikšmingumo  $p$  – vertės. Testas atliekamas su statistiniais įverčiais apibrėžiamais formule:

$$W = \left| \sum_{i=1}^{N_r} [\text{sgn}(x_{2,i} - x_{1,i}) \cdot R_i] \right|, \quad (17)$$

čia  $R_i$  –  $i$  – osios poros statistinis rangas kaip atskirų rangų vidurkis (angl. statistical ranking), gaunamas kaip indeksas atitinkantis vietą surūšiuotuose duomenyse nuo mažiausio iki didžiausio.

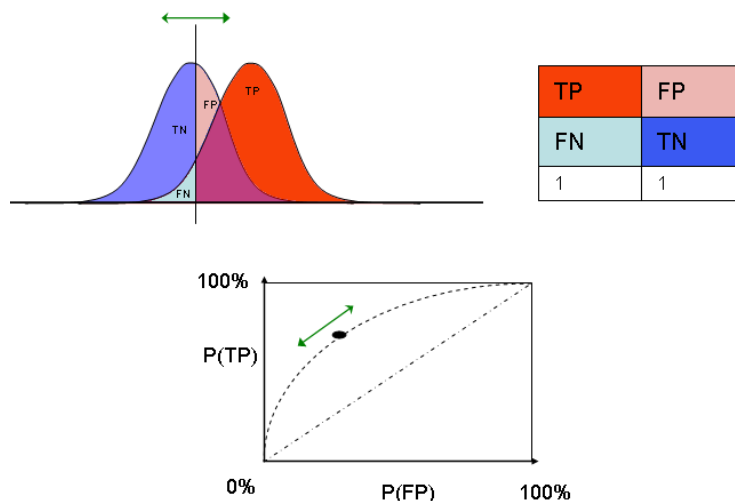
Šis metodas išsiskiria iš Bhattacharyya atstumo, santykinės entropijos ir Stjudento  $t$  – testo dėl to jog nereikalauja prielaidos apie normalinį duomenų pasiskirstymą ir yra neparametrinis, tai pagrindinis jo pranašumas prieš šiuos metodus.

Metodo trūkumai tie patys kaip ir pastarųjų metodų: prielaida apie duomenų nepriklausomumą.

### 2.4.5. ROC charakteristika

ROC (angl. Receiver Operating Characteristic) kreivė nusakomas duomenų imčių poros persidengimas. Plotas po ROC kreivė nusako kiek duomenys yra atskiriami: kuo plotas didesnis tuo duomenų poros imtys yra labiau atsiskyrusios. Pagal tai ir yra indeksuojami požymiai. Kaip ir prieš tai minėti metodai šis metodas taip pat yra porinis, o esant daugiau nei kelioms klasėms imamas porų indeksų vidurkis kaip skiriamumo indeksas. Šia kreivė nurodomas binarinio tiesinio klasifikatoriaus tikslumas keičiant slenkstį visame intervale.

ROC charakteristika iliustruota 2.4.5.1 paveiksle: turint du skirstinius yra sudaroma krevė iš šių skirstinių integralų (angl. cumulative density function) vienoje ar kitoje pusėje nuo slenkščio vietos, t.y., grafikas sudaromas kaip dviejų klasių tikimybių skirstinių plotų iki slenkščio priklausomybės tarp teigiamos klasės (angl. true positive) ir neteisingai priskirtos neigiamos klasės (angl. false positive) funkcija keičiant slenkstį visame intervale. Įžambinė grafike roto tą atvejį kai duomenys yra visiškai persidengę, o plotas skaičiuojamas tarp šios įžambinės ir gautos ROC kreivės.



2.4.5.1 pav. ROC charakteristikos gavimo pavyzdžio iliustracija. Maksimalaus nuotolio nuo įžambinės taškas rodo optimaliausią binarinio klasifikatoriaus slenkščio vietą



Kaip ir Wilcoxon testas, šis metodas yra neparametrinis ir papildomai nei Wilcoxon testas, jis nereikalauja jokios prielaidos apie duomenų pasiskirstymų vienodumą poroje. Tai šio metodo privalumas lyginant su kitais metodais.

Trūkumas tas jog kaip ir visi kiti metodai šis metodas priima prielaidą apie duomenų tarpusavio nepriklausomumą.

#### 2.4.6. RELIEFF požymių atrinkimo algoritmas

Šio algoritmo pagrindinis privalumas jog jis nėra priklausomas nuo prielaidų kaip kiti metodai. Jis taip pat tinkamas ir tuo atveju kai klasių skaičius didesnis nei dvi. Pirmoji algoritmo versija buvo pavadinta RELIEF ir naudojama dviejų klasių atveju:

- Turima įėjimų matricą su  $n$  stebėjimų bei  $p$  požymių kiekiais, priklausančiais dviem žinomoms klasėms;
- Kiekvieno požymio vektoriaus stebėjimai yra normalizuojami iki  $[0, 1]$  intervalo pločio, o klasėms priskiriamos binarinės vertės (0 arba 1);
- Algoritmas bus kartojamas  $m$  kartų, pradedant su  $p$  ilgio nulinių svorių vektoriumi ( $W$ );
- Kiekvienos iteracijos metu, atsitiktinai paimamas vieno stebėjimo požymių vektorius ( $X$ ) ir kiekvienos klasės požymių vektoriai artimiausi pirmajam atsitiktinai parinktam vektoriui pagal nurodytą atstumą (pvz. Euklido). Artimiausi tos pačios klasės stebėjimai vadinami „near – hit“, o artimiausi skirtingų klasių stebėjimai – „near – miss“.
- Sviurių vektorius atnaujinamas pagal tokią formulę:

$$W_i = W_{i-1} - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 . \quad (18)$$

- Kiekvienu atveju požymio svorio vertė sumažėja, jei atstumas iki artimiausio stebėjimo (pagal pasirinktą atstumą) kito taško („near – hit“) toje pačioje klasėje yra didesnis nei artimiausio kitos klasės stebėjimo taško („near – miss“). O svoriai padidinami jei būna atvirkščiai.
- Po  $m$  iteracijų kiekvienas svorių vektoriau selementas padalinamas iš  $m$ .

Nauja šio algoritmo versija pavadinta RELIEFF buvo patobulinta senojo (aukščiau aprašyto) algoritmo versija:

- Vietoj to jog algoritmas kartojamas  $m$  kartų, jis buvo pritaikytas kaip pilnas perrinkimas ( $n$  kartų), kai  $n$  palyginti mažas (iki 1000);

- Vietoj to jog ieškomas vienas artimiausias taškas, buvo nuspręsta naudoti  $k$  artimiausių kaimynų paiešką, užtikrinant labiau robastišką sprendimą;
- Algoritmas tapo pritaikomas kai klasių skaičius didesnis nei dvi, RELIEFF ieško  $k$  „near-miss“ verčių iš kiekvienos klasės ir paima šių klasių indėlio vidurkį  $W$  vektoriaus atnaujinimui.

#### 2.4.7. Atsitiktinio miško požymių indeksai

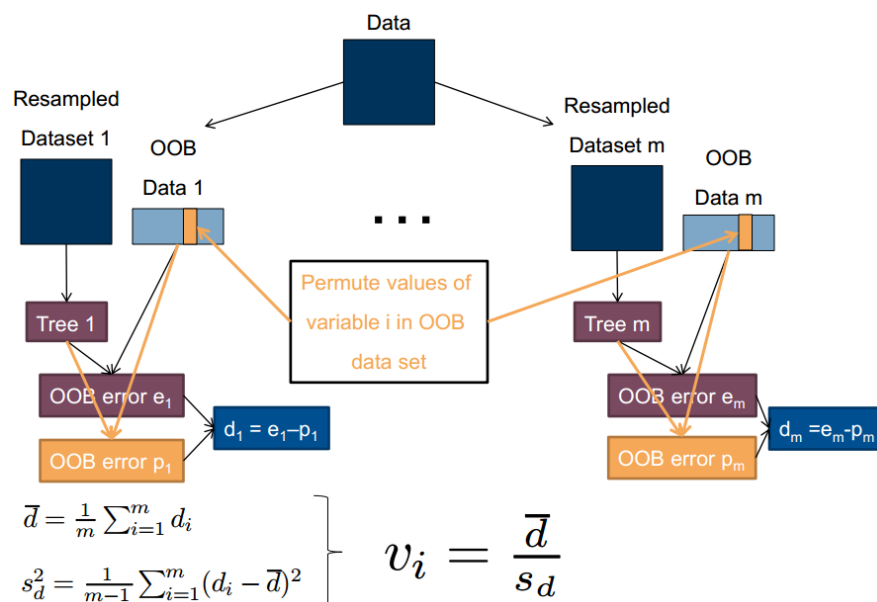
Nors atsitiktinis miškas (angl. Random Forest) yra klasifikavimo algoritmas, jis gali būti panaudojamas kaip požymių atrankos metodas. Toliau aprašomas atsitiktinio miško pritaikymas požymiams atrinkti:

- Tarkime, yra nustatomas  $X^j$  požymio vektoriaus svarbumo indeksas. Kiekvienam klasifikavimo medžiui  $t$  yra išskirtas  $OOB_t$  (angl. Out – Of – Bag) duomenų imtis, kuri nebuvo įtraukta į atitinkamo klasifikavimo medžio  $t$  apmokymo imtį. Atitinkamo medžio klasifikavimo klaida  $errOOB_t$  apibūdinama kaip klasifikavimo klaida nustatyta su  $OOB_t$  duomenų testavimo imtimi;
- Toliau, atsitiktinai pakeičiamos  $X^j$  požymio vektoriaus vertės  $OOB_t$  duomenų imtyje. Gaunama sukeista duomenų imtis  $\widehat{OOB}_t^j$  ir skaičiuojama klaida  $err\widehat{OOB}_t^j$  (klasifikavimo klaida su  $\widehat{OOB}_t^j$  ant medžio  $t$ ).
- Požymio svarbos indeksas nusakomas pagal formulę:
- 

$$VI(X^j) = \frac{1}{ntree} \sum_t (err\widehat{OOB}_t^j - errOOB_t), \quad (19)$$

čia  $ntree$  – klasifikavimo medžių skaičius atsitiktiniame miške.

Požymių indeksavimo algoritmo iliustracija atsitiktiniame miške pavaizduota 2.4.7.1 paveiksle.

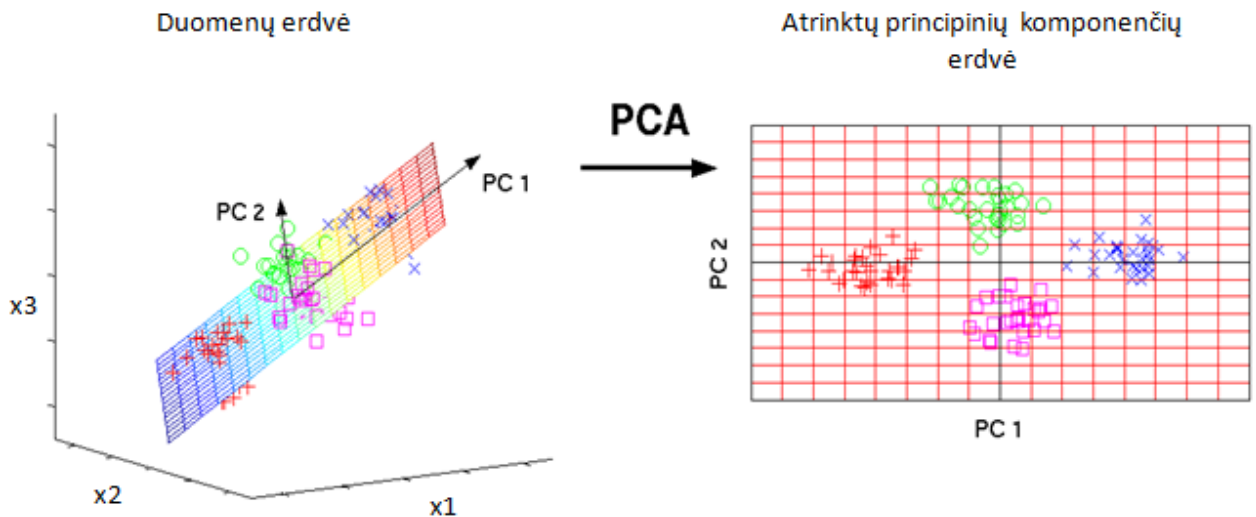


2.4.7.1 pav. Požymių svarbumo indeksavimo algoritmo iliustracija

## 2.4.8. Požymių transformavimas naudojant PCA

Šiame skyryje aptariamas principinių komponentių analizės metodas taikomas klasifikavimo požymiams gauti.

Principinės komponentės naudojamos požymių dimensiškumui sumažinti išrenkant tam tikrą kiekį duomenų principinių komponentių, kurių kiekis mažesnis nei duomenų požymių kiekis ir jos gaunamos kaip tiesinė įėjimo požymių kombinacija. Principinės komponentės išgaunamos naudojant duomenų bazinius vektorius (angl. eigenvectors) ir jų bazines vertes (angl. eigenvalues). Baziniai vektoriai – ortogonalūs vektoriai kurių kryptys nustatomos pagal maksimalią variaciją duomenyse – bazines vertes. Principinių komponentių pagalba baziniai vektoriai surūšiuojami pagal jų bazines vertes priimant prielaidą kad duomenis geriausia analizuoti (klasifikuoti) pagal maksimalios duomenų variacijos kryptį. Atrenkamos pirmos kelios principinės komponentės (kelių bazinių vektorių bei originalių duomenų vektorinė sandauga) kurios naudojamos tolesnei analizei. Principinių komponentių analizė grafiškai iliustruota 2.4.8.1 pav.



2.4.8.1 pav. Duomenų dimensiškumo sumažinimo naudojant principinių komponentių analizę pavyzdys

## 2.5. Klasifikavimo rezultatai

Toliau aptariamai gauti klasifikavimo rezultatai su įvairiais požymių rinkiniais gautais atlikus požymių rangvimą taikant anksčiau aprašytus metodus. 2.5.1 lentelėje pateikiami požymių aprašymai ir jų eilės numeriai, kurie bus toliau naudojami apibūdinant požymių rinkinius.

2.5.1 lentelė. Požymių pavadinimai ir juos atitinkantys numeriai

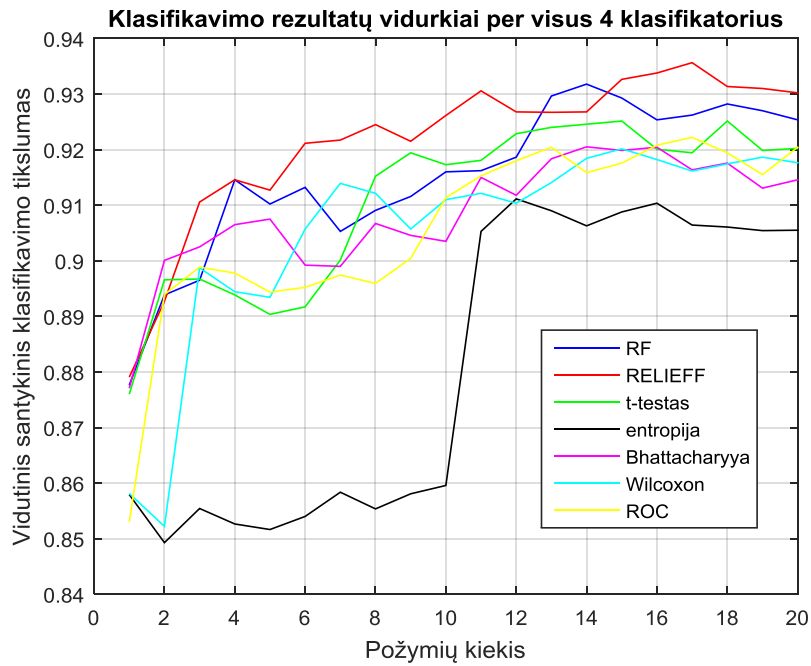
Požymio pavadinimas	Eilės numeris
Y ašies signalo kvantiliai (0,1:0,1:0,9)	1,4,7,10,13,16,19,22,25
Z ašies signalo kvantiliai (0,1:0,1:0,9)	2,5,8,11,14,17,20,23,26
X signalo kvantiliai (0,1:0,1:0,9)	3,6,9,12,15,18,21,24,27
Atstumas tarp Y ašies signalo kvantilių (0,1-0,9;0,2-0,8;0,3-0,7;0,4-0,6)	28,31,34,37
Atstumas tarp Z ašies signalo kvantilių (0,1-0,9;0,2-0,8;0,3-0,7;0,4-0,6)	29,32,35,38
Atstumas tarp X ašies signalo kvantilių (0,1-0,9;0,2-0,8;0,3-0,7;0,4-0,6)	30,33,36,39
Koreliacijos koef. tarp Y ir Z ašių signalų	40
Koreliacijos koef. tarp Z ir X ašių signalų	41
Koreliacijos koef. tarp Y ir X ašių signalų	42
Y ašies signalo duomenų ribos viršijimo intensyvumas	43
Z ašies signalo duomenų ribos viršijimo intensyvumas	44
X ašies signalo duomenų ribos viršijimo intensyvumas	45
Y ašies signalo trūkių standartinis nuokrypis	46
Z ašies signalo trūkių standartinis nuokrypis	47
X ašies signalo trūkių standartinis nuokrypis	48
Y ašies signalo santykinė spektro galia (0,1:0,1:1 norm. diapazonuose)	49:58
Z ašies signalo santykinė spektro galia (0,1:0,1:1 norm. diapazonuose)	59:68
X ašies signalo santykinė spektro galia (0,1:0,1:1 norm. diapazonuose)	69:78

Klasifikavimas buvo atliktas pridedant kas vieną nuo geriausio iki prasčiausio pagal požymio rangą rūšiuotą požymį. Iš visų 78 požymių buvo pasirinkti 20 pirmų ranguočių požymių pagal kiekvieną iš požymių atrinkimo metodų. 2.5.2 lentelėje pateikti rūšiuoti požymių numeriai nuo geriausio iki prasčiausio atliekant rangavimą su atitinkamu metodu.

2.5.2 lentelė. Požymių rangavimo rezultatai su skirtingais metodais

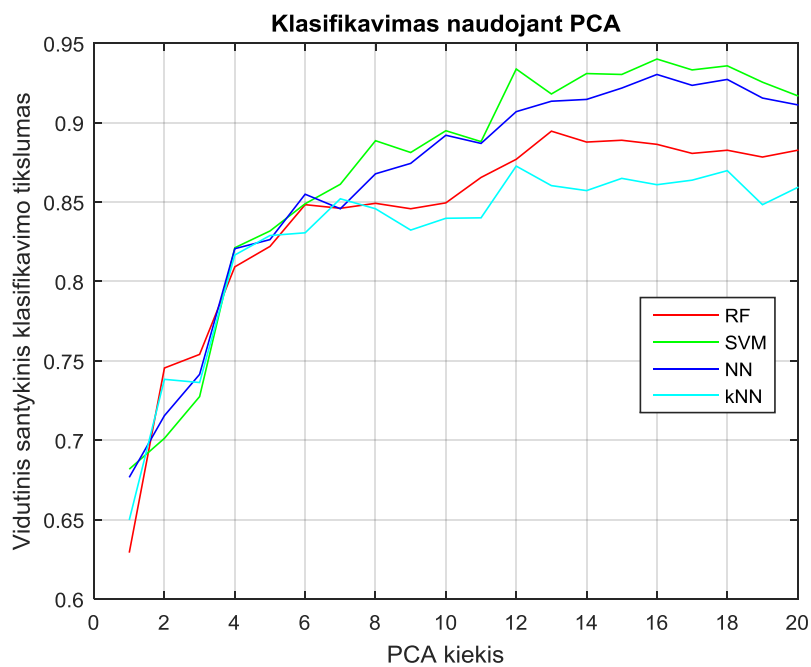
Eilės numeris (1-geriausias)	Rangavimas pagal atsitiktinių miškų (RF) indeksus	Rangavimas pagal RELIEFF indeksus	Rangavimas pagal Studento t-testo indeksus	Rangavimas pagal santykinės entropijos indeksus	Rangavimas pagal Bhattacharyya indeksus	Rangavimas pagal Wilcoxon indeksus	Rangavimas pagal ROC indeksus
1	48	48	48	69	48	69	69
2	69	45	69	74	69	70	48
3	45	3	45	70	74	48	45
4	3	78	78	76	70	45	74
5	74	69	70	71	76	74	70
6	6	30	74	64	45	3	76
7	70	6	76	61	77	71	77
8	78	25	3	77	75	76	78
9	76	43	30	49	71	77	75
10	25	41	75	67	78	6	30
11	9	33	77	48	3	78	3
12	77	9	25	75	49	49	71
13	17	27	33	59	30	75	25
14	14	22	6	65	25	30	43
15	75	44	22	60	56	51	33
16	22	26	43	56	61	9	22
17	12	36	36	51	64	2	6
18	30	23	71	68	67	25	36
19	71	70	19	55	54	5	49
20	73	76	39	54	51	8	41

Toliau aptariami gauti agresyvaus ir saugaus vairavimo stiliaus klasifikavimo rezultatai. 2.5.1 paveiksle pateikti vidutiniai klasifikavimo rezultatai (per visus klasifikatorius), kada pridedama po vieną ranguočių požymį kiekvieno rangavimo metodo atvejui. Kaip matyti iš grafiku, vidutiniškai didžiausią klasifikavimo tikslumą galima pasiekti naudojant pagal RELIEFF algoritmą ranguotus požymius. Kaip matyti jau su pirmais 3 rūšiuotais (pagal RELIEFF) požymiais pasiekiamas ~91% klasifikavimo tikslumas.

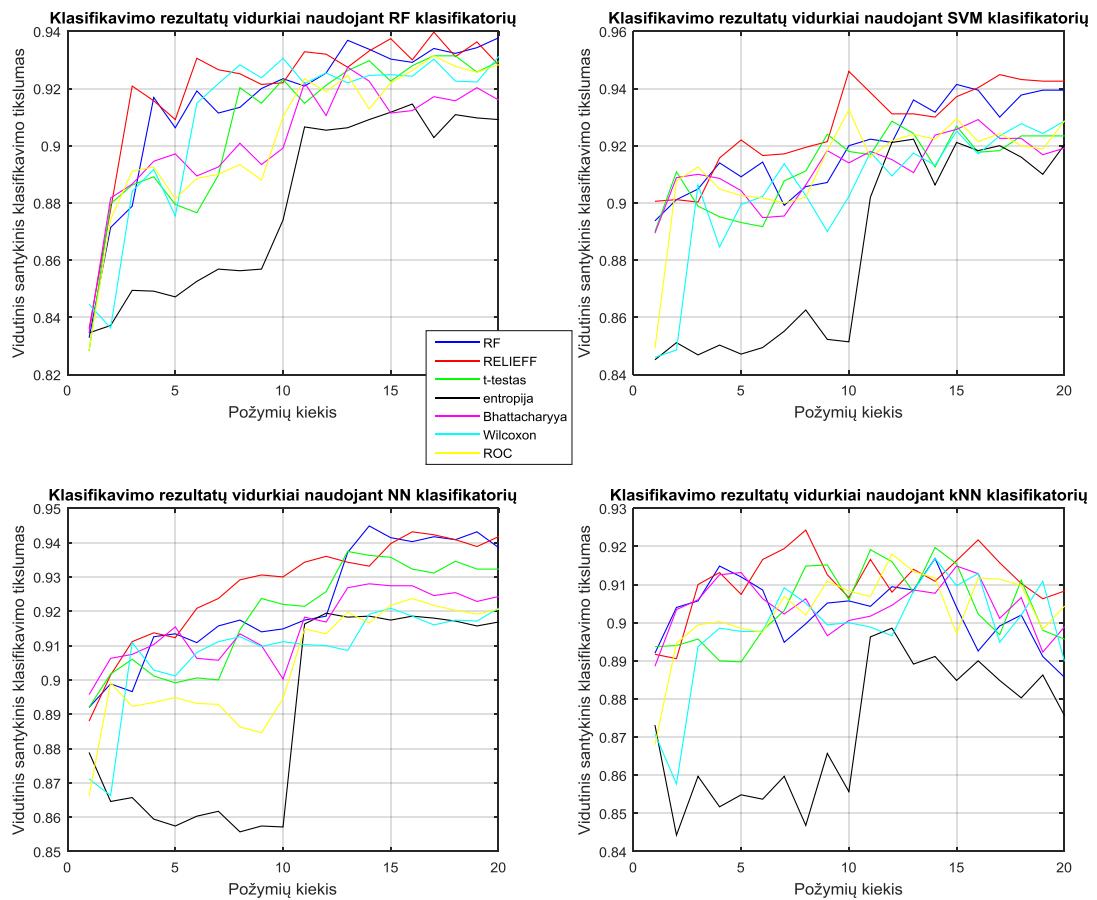


2.5.1 pav. Klasifikavimo rezultatų vidurkiai (per visus klasifikatorius) pridedant po vieną ranguotą požymį

Rezultatai kiekvienam klasifikatoriui atskirai pateikti 2.5.2 (PCA atvejais), bei 2.5.3 (visi kiti atvejai) paveiksluose. Iš 2.5.2 pav. matyti jog naudojant PCA požymius didžiausias vidutinis klasifikavimo tikslumas yra  $\sim 95\%$  su atraminių vektorių klasifikatoriumi. Iš 2.5.3 pav. pateiktų grafikų matyti, kad RELIEFF požymių naudojimas (kai kuriais atvejais ir RF požymių naudojimas) leidžia pasiekti geriausias rezultatus su kiekvienu iš klasifikatorių.



2.5.2 pav. Klasifikavimo rezultatai naudojant PCA transformuotus požymius kiekvienam klasifikatoriui atskirai



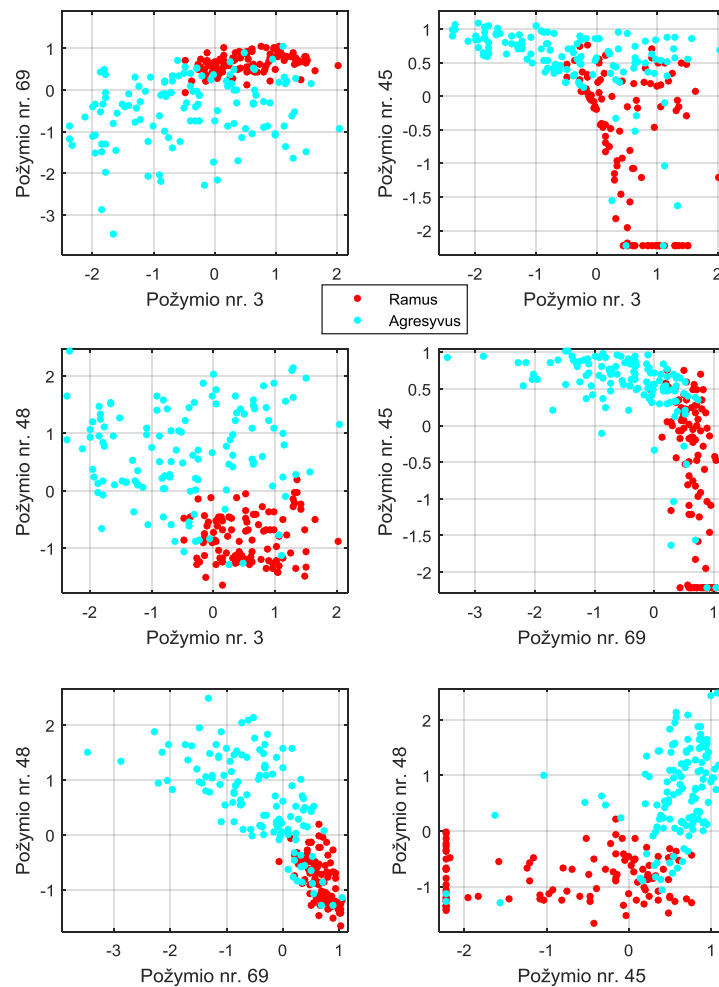
2.5.3 pav. Klasifikavimo rezultatai naudojant ranguotus požymius kiekvienam klasifikatoriui atskirai

2.5.3 lentelėje palyginami rezultatai kada naudojami visi požymiai bei kada naudojamas požymių atrinkimas/transformavimas. Kaip matyti visais atvejais požymių atrinkimas pagerino agresyvaus ir saugaus vairavimo klasifikavimų rezultatus, o didžiausias tikslumas pasiektas naudojant atraminių vektorių klasifikavimo algoritmą.

2.5.3 lentelė. Klasifikavimo tikslumų palyginimas naudojant visus ir atrinktus požymius

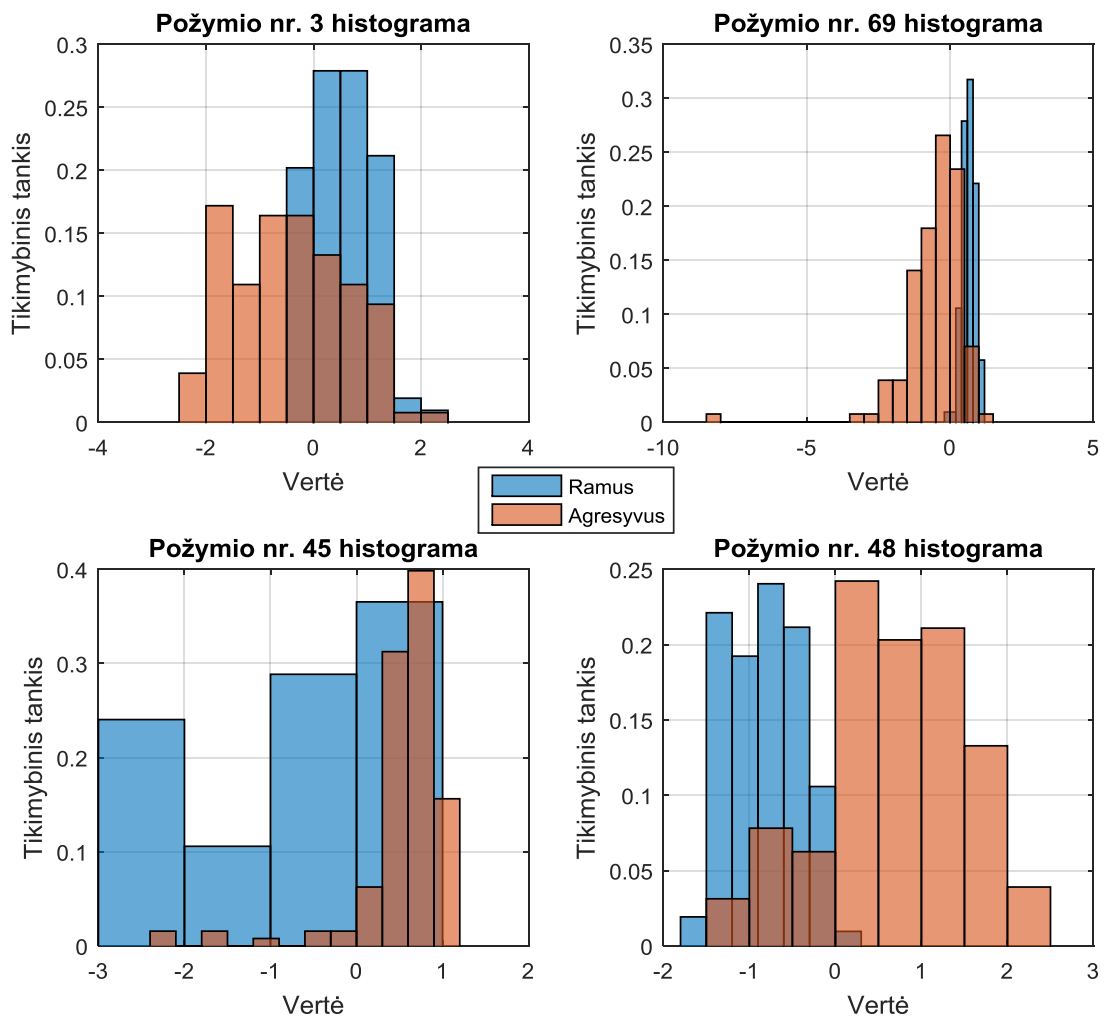
Klasifikatoriaus pavadinimas	kNN	SVM	NN	RF
Naudojant visus požymius	85,771%	92,743%	92,600%	92,743%
Naudojant geriausią požymių rinkinį	92,429%	94,600%	94,486%	93,971%

Siekiant atvaizduoti klasių pasiskirstymą požymių erdvėje parinkti 4 pirmi rūšiuoti požymiai, kuriuos naudojant pasiekiamas ~91,5% vidutinis klasifikavimo tikslumas per visus klasifikatorius. Agresyvaus ir saugaus vairavimo pasiskirstymai atvaizduojant požymius poromis pateikti 2.5.4 paveiksle, o kiekvieno iš požymių histogramas pateiktos 2.5.5 pav.



2.5.4 pav. Požymių pasiskirstymo į agresyvias ir saugias klases vizualizavimas poromis





2.5.5 pav. 4 geriausių požymių histogramos kiekvienai klasei

Išsamiau paanalizavus 4 geriausius požymius matyti, jog jie visi išgauti tik iš X ašies G – jutiklio signalo (žr. 1 lentelę):

- požymis nr. 3 yra 0,1 X ašies signalo kvantilis. Kaip iliustruota 4 pav. agresyvaus vairavimo kvantilis turi mažesnę reikšmę nei saugaus vairavimo. Tai susiję su greitu automobilio startavimu iš vietos ir staigiu įsibėjėjimu.
- požymis nr. 45 rodo X ašies signalo duomenų ribos viršijimo intensyvumą padauginus iš viršijimo intensyvumo amplitudės. Kaip požymis paimtas šios vertės logaritmas. Tai susiję su staigiu transporto priemonės stabdymu ar akseleravimu, kada reikšmingai viršijama iš saugaus vairavimo nustatyta signalo vertės riba (žr. 7 paveikslą).
- požymis nr. 48 rodo trūkių standartinę nuokrypį X ašies signale. Šis požymis atrinktas kaip pats geriausias požymis ir pats vienas leidžia pasiekti iki ~89% su SVM klasifikatoriumi. Didelės trūkių vertės yra tiesiogiai susijusios su keleivio diskomfortu, didesnėmis kuro sąnaudomis ir vairavimo agresyvumu (žr. 8 paveikslą).

- požymis nr. 69 rodo vidutinį galios spektro intensyvumą 0-0,1 dažnių juostos diapazone (dažnių juosta normuota iki 1). Kaip buvo atvaizduota 2 paveiksle, matyti kad agresyvus vairavimas turi didesnę galią šiame diapazone. Šis diapazonas rodo aukštų dažnių harmonikų buvimą signale kai vairavimas agresyvus (siauri trumpalaikiai X ašies pagreičio signalo pakilimai ir nusileidimai).

### **3. VIEŠOJO TRANSPORTO AUTOBUSŲ VAIRUOTOJŲ ELGSENOS KLASIFIKAVIMAS**

Vairuotojo elgsenos klasifikavimas viešojo transporto sektoriuje gali padėti identifikuoti vairuotojo profesionalumą: žinant etaloninį profesionalų vairuotoją galima lyginti kitus vairuotojus kiek arti jie yra. Pagal vieno vairuotojo panašumą į kitą galima nustatyti jo profesionalumą arba kitus etaloniniam vairuotojui būdingos vairavimo elgsenos požymius. Taip pat naudojant požymius susijusius su agresyviu vairavimu (kaip aprašyta ankstesnio agresyvaus ir saugaus vairavimo klasifikavimo tyrimo metu) galima nustatyti vairuotojo agresyvumą (lyginimo principu, analogišku vairuotojo profesionalumo nustatymui).

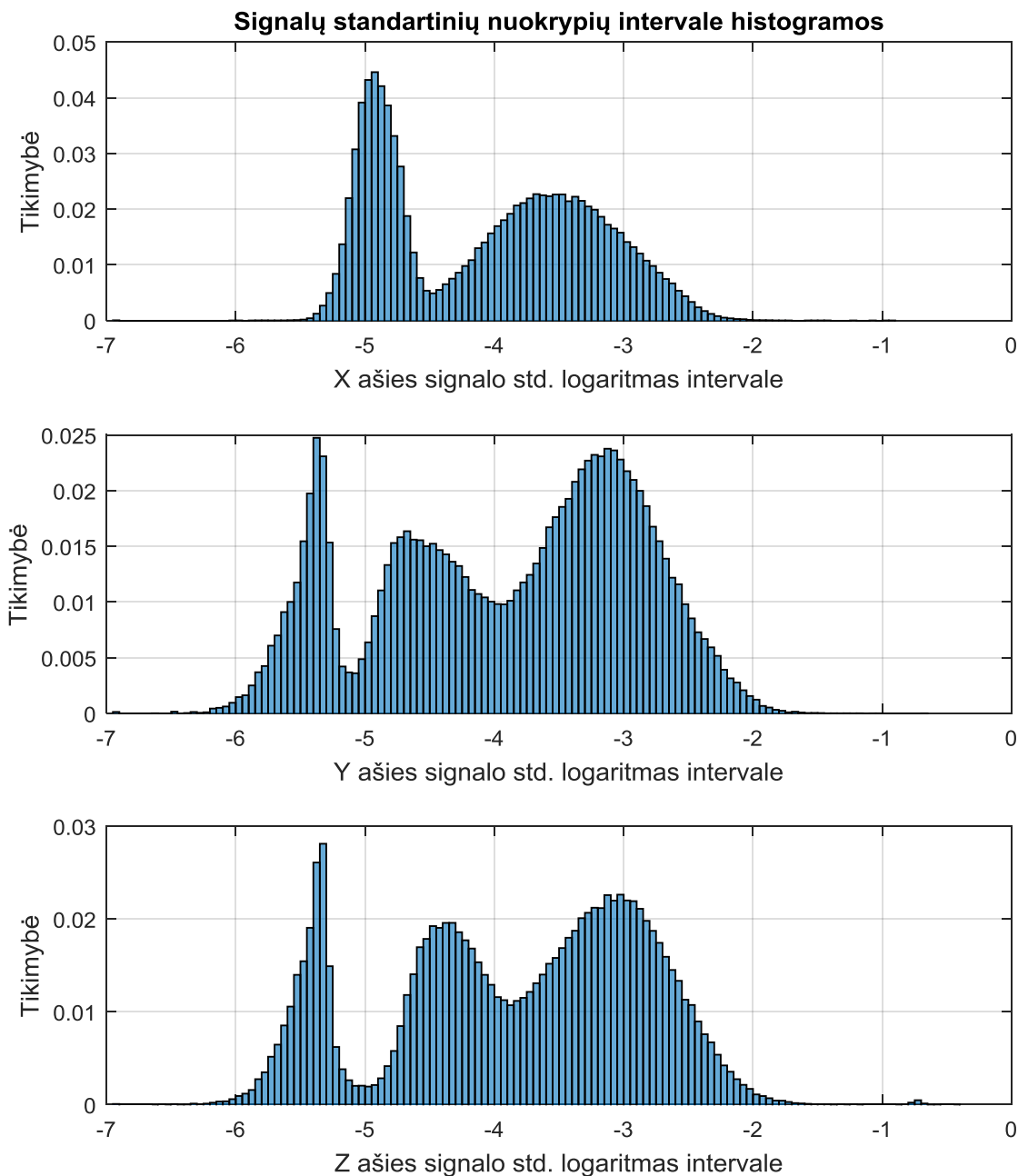
Toliau šiame skyriuje panaudojami anksčiau aprašyti požymiai, kurie taip pat išgaunami iš trijų ašių akselerometro signalų, pritaikomi viešojo transporto autobusų vairuotojų klasifikavimui. Eksperimentiniai duomenys naudojami šiame tyrime surinkti iš važiavimų vienu viešojo transporto autobusu dviems vairuotojams 12 savaitių laikotarpyje. Signalai surinkti tuo pačiu 17 Hz dažniu. Važiavimai atlikti įvairiais pasikartojančiais maršrutais mieste vairuotojams dažniausiai dirbant pamainomis (vienas vairuotojas veža keleivius pirmą dienos pusę, o kitas – antrą, bei atvirkščiai).

#### **3.1. Signalų apdorojimas**

Signalų apdorojimas atliekamas analogiškas aprašytam 2.1 skyriuje: signalo išskirčių pašalinimas ir triukšmo nufiltravimas. Bet papildomai nei prieš tai nagrinėtu atveju, realūs signalai surenkami važiuojant viešojo transporto autobusui yra pilni pašalinės sustojimo informacijos: signalas yra generuojamas nors tiriamoji transporto priemonė stovi ir negeneruoja jokios svarbios informacijos apie vairavimo elgseną. Tam reikalinga sumažinti (idealiu atveju eliminuoti) sustojimų informaciją surenkamuose signaluose.

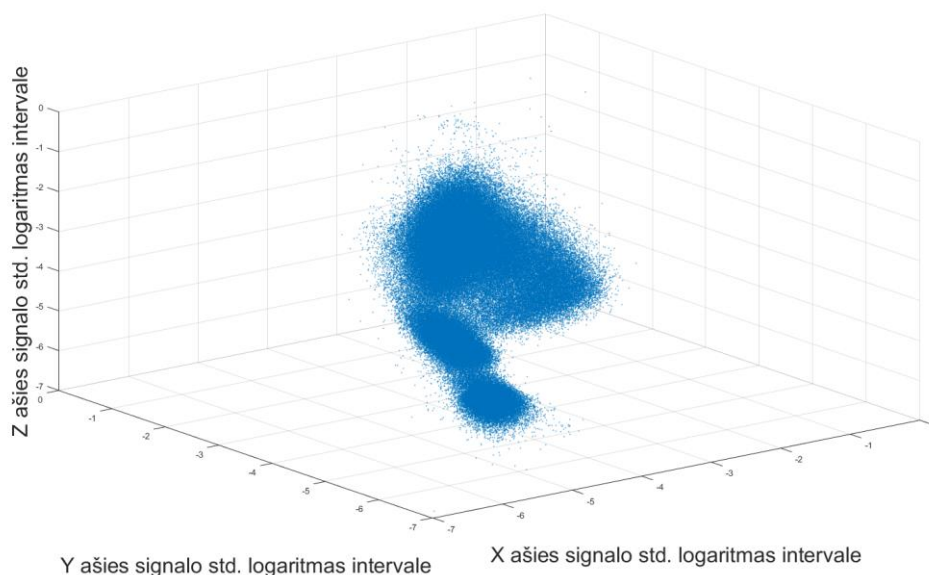
Tam tikslui nagrinėsime signalo standartinius nuokrypius siaurame 100 diskretų (~6 sekundžių) intervale. Standartiniai nuokrypiai koduoja informaciją apie važiavimo signalo buvimą arba nebuvimą: didenis standartinis nuokrypis rodo jog vyksta važiavimas, o mažas – jog autobusas stovi. Pagrindinė problema tokiu atveju yra optimalios standartinio nuokrypio ribos  $G$

– jutiklio signaluose nustatymas. 3.1.1 paveiksle pateiktos standartinių nuokrypių intervale histogramos kiekvienam signalui. Iš jų matyti jog signalus sudaro kelios tikimybinio tankio funkcijos su skirtingais centrais. X ašies atveju matyti ryškus bimodalinis skirstinys, o kituose ašyse – skirstiniai su trimis ryškiomis tikimybių tankių funkcijomis.



3.1.1 pav. G – jutiklio signalų standartinių nuokrypių 100 diskretų intervaluose logaritmų skirstiniai

Tik viena šiuo atveju tankio funkcija yra susijusi su važiavimu, dvi (Y ir Z ašies atveju) ir viena (X ašies atveju) yra susijusios su duomenų generavimu stovėjimo metu. Siekiant apjungti trijų signalų informaciją trijų dimensijų vizualizacija pateikta 3.1.2 pav. Pažvelgus į 3.1.2 pav. matyti jog gaunami aiškūs stovėjimo klasteriai. Stovėjimo duomenų klasteriai yra du: 1) kai transporto priemonės variklis nedirba; 2) kai transporto priemonės variklis dirba, bet ji stovi. Antruoju atveju gaunami didesni standartiniai nuokrypiai, bet jie išsiskiria nuo važiavimo duomenų standartinių nuokrypių.



3.1.2 pav. Trijų dimensijų G – jutiklio signalų standartinių nuokrypių logaritmų intervale klasterių išsidėstymų vizualizacija

Remiantis informacija matoma iš 3.1.1 ir 3.1.2 paveikslų atrodytų patogu pritaikyti Gauso mišinių modelių klasterizavimo algoritmą automatiniam stovėjimo duomenų identifikavimui.

Gauso mišinių modelis (angl. Gaussian Mixture Model (GMM)) apibrėžiamas kaip parametrinė tikimybės tankio funkcija išreikšta tiesine kombinacija pasvertų multidimensinių Gauso pasiskirstymo funkcijų:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^M w_i g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad \lambda = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}, \quad (20)$$

čia  $\mathbf{x}$  –  $D$  – dimensinis duomenų vektorius (požymių matrica);

$M$  – Gauso funkcijų kiekis;

$w_i$  – atitinkamos Gauso funkcijos svoris (visų mišinių svorių suma lygi 1);

$g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  – Gauso funkcija su vidurkio ir dispersijų multidimensinėmis matricomis.

Daugadimensė Gauso tankio funkcija apibrėžiama formule:

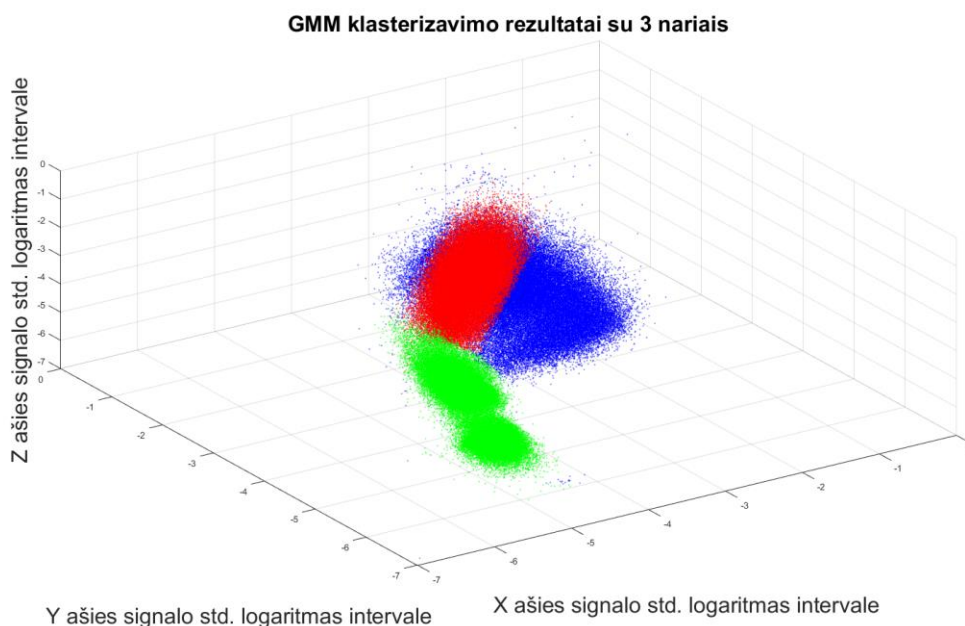
$$g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}, \quad (21)$$

čia  $\boldsymbol{\mu}_i$  – multidimensinis  $i$  – osios Gauso funkcijos vidurkio vektorius;

$\boldsymbol{\Sigma}_i$  –  $i$  – osios Gauso funkcijos kovariacijos matrica.

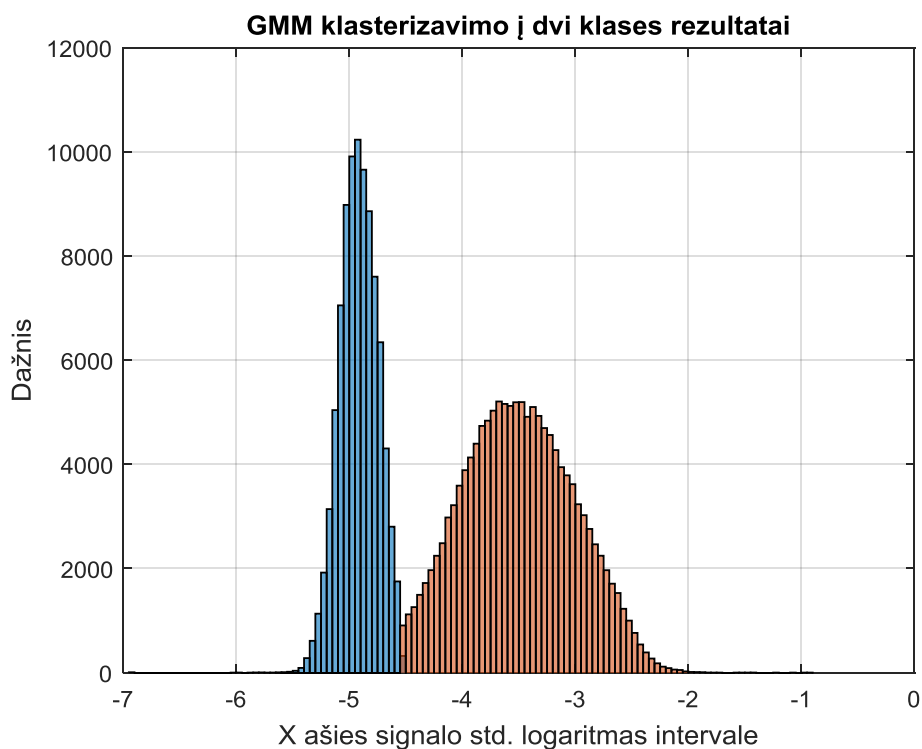
Gauso mišinių (mikstūrų) taikymas klasterizavimui yra susijęs su tikėtinumo – maksimizavimo algoritmu (angl. Expectation – Maximization). Tai iteracinis algoritmas naudojamas Gauso mišinių pasiskirstymo modelio parametrų pritaikymui duomenų segmentams. Duomenų taškas priskiriamas tam tikram segmentui pagal didžiausią tikimybės vertę. Šio modelio privalumas, jog duomenų taško priklausymas vienam ar kitam segmentui išreiškiamas per tikimybes.

Kadangi stovėjimui būdingos dvi tikimybių funkcijos, priimama prielaida jog trijų tikimybių funkcijų GMM modelis turėtų automatiškai priskirti į klasterius stovėjimo duomenis kai autobuso variklis nedirba, stovėjimo duomenis kai autobuso variklis dirba, bei važiavimo duomenis. GMM modelio taikymas klasterizavimui naudojant pilną trijų dimensijų erdvę pateiktas 3.1.3 pav. Kaip matyti klasterizavimo rezultatai netenkina: nors abu stovėjimų duomenų klasteriai priskirti į vieną (žalią) klasterį, kiti klasteriai suformuojami važiavimo duomenyse, kas rodo jog trijų narių klasterizavimas nėra tinkamas ir gali būti rizikingas kitais atvejais (dėl skirtingos įjungto variklio vibracijos įtakos stovėjimo duomenims).



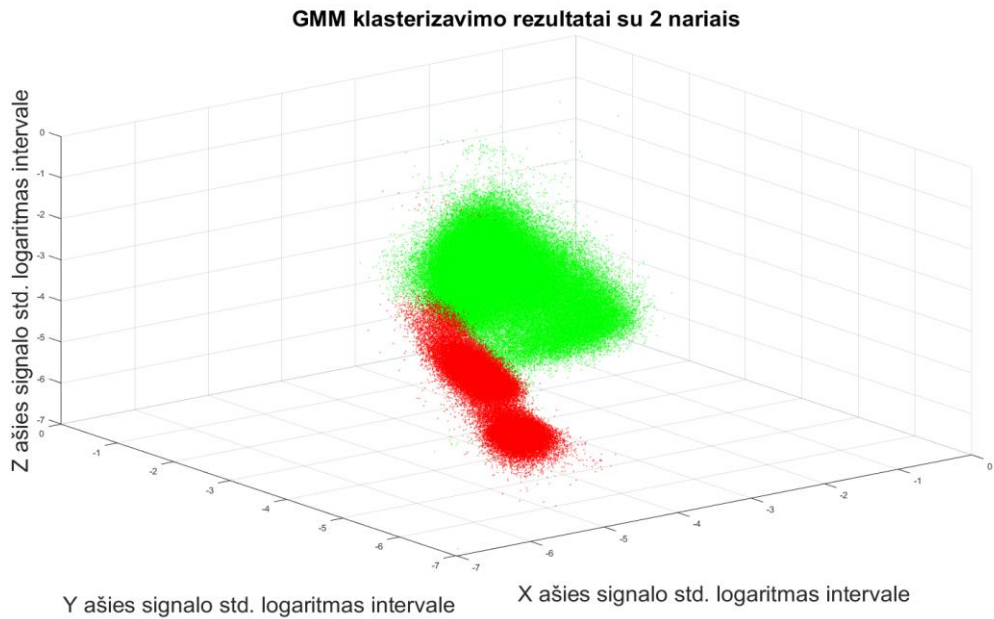
3.1.3 pav. GMM klasterizavimas naudojant pilną signalų erdvę su trimis nariais

Kadangi iš 3.1.1 pav. buvo matyti X ašies signalo ryškus pasiskirstymas į dvi Gauso funkcijas, GMM modelis pritaikomas vienos ašies (X) signalui klasterizuoti (optimaliai ribai nustatyti). Gauta optimali skiriamoji riba bus naudojama stovėjimo signalams atskirti nuo važiavimo signalų. 3.1.4 pav. pateiktas GMM modelio taikymas su dviem nariais.



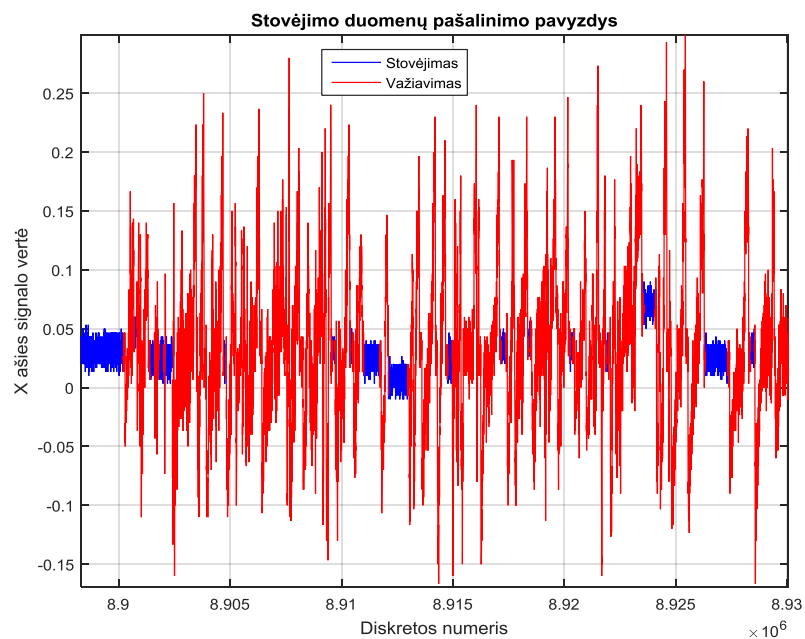
3.1.4 pav. GMM modelio su dviem nariais X ašies klasterizavimo rezultatai

Kaip matyti automatiškai pavyksta nustatyti optimalią skiriamąją ribą, o šios ribos panaudojimas trijų dimensijų duomenų erdvėje pateiktas 3.1.5 pav. Matoma, jog stovėjimai kokybiškai atskiriami nuo važiavimo duomenų. Kadangi panaudojama tik X ašies informacija, tai riba galimai atriboja važiavimo duomenų dalį į stovėjimo duomenis. Šiuo atveju tikslumas nėra tiek svarbus, svarbu sumažinti stovėjimo duomenų įtaką važiavimo informacijoje.



3.1.5 pav. GMM modelio optimalios skiriamosios ribos panaudojimas trijų dimensijų duomenims atskirti

3.1.6 pav. pateiktas autobuso X ašies važiavimo signalas po stovėjimo duomenų pašalinimo taikant vienos ašies (X) GMM modeliavimą. Kaip matyti stovėjimo duomenys su ilgu stovėjimo laikotarpiu pašalinami kokybiškai. Taip pat matyti, jog paprasto slenksčio taikymas duomenims pašalinti būtų neefektyvus, nes stovėjimas vyksta įvairuose signalo lygiuose.



3.1.6 pav. X ašies signalas po stovėjimo duomenų pašalinimo taikant vienos ašies GMM modeliavimą.

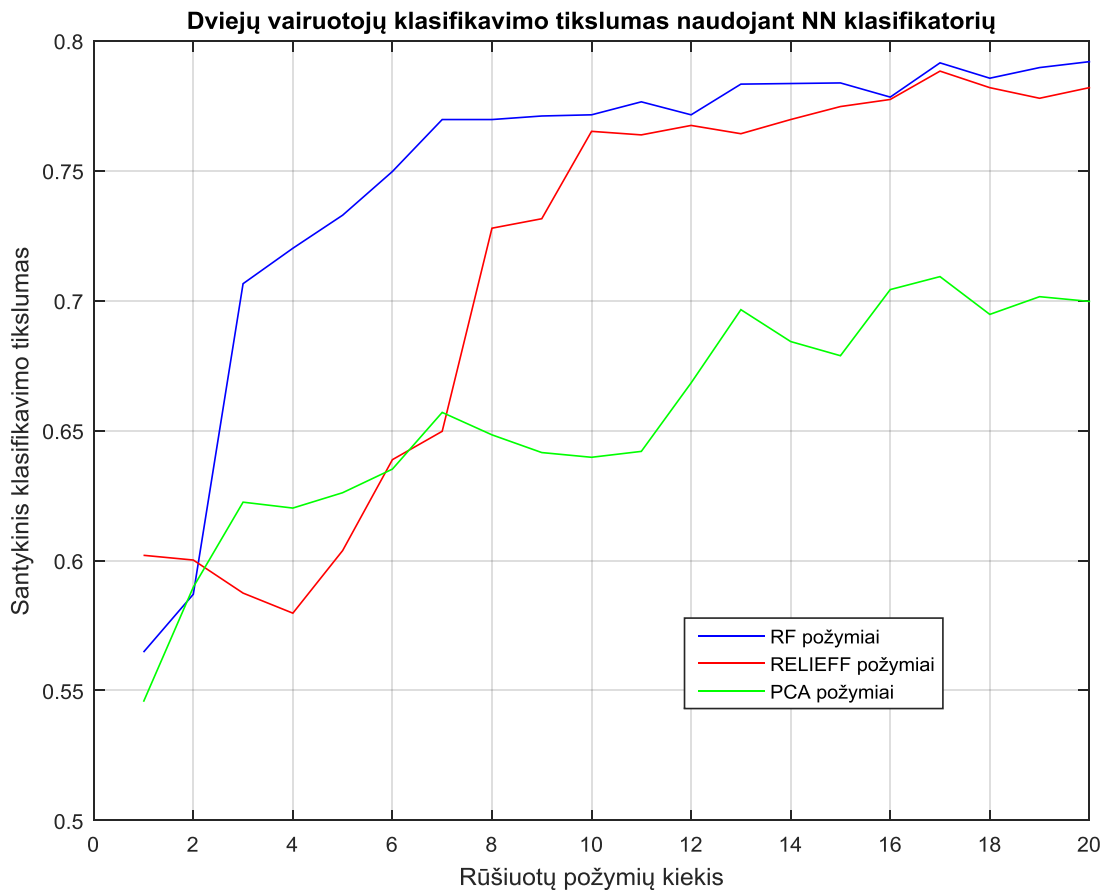
### 3.2. Vairavimo elgsenos klasifikavimas

Toliau šiame poskyriuje aptariami dviejų autobuso vairuotojų elgsenos klasifikavimo rezultatai. Klasifikavimas buvo atliekamas išgaunant požymius iš 3000 diskretų intervalo važiavimo duomenų (po filtravimo ir stovėjimų pašalinimo). Požymiai išgaunami analogiškai kaip ir vairuotojo agresyvumo klasifikavimo atveju. Taigi, kadangi šiuo atveju vairuotojų klasifikavimo tyrimui naudoti tie patys požymiai, vairuotojų tarpusavio atsiskiriamumas rodys jog vienas vairuotojas yra daugiau ar mažiau agresyvesnis už kitą. Tyrimo objektas šiuo atveju yra vairuotojų elgsenos klasifikavimas, kurią apibūdina su agresyvumu susiję požymiai, bet tai nėra tapatu agresyviai ar saugiam vairavimo klasifikavimo uždaviniui, kadangi kaip žymės (angl. labels) naudojamos duomenų taško priklausymas vienam ar kitam vairuotojui, o ne agresyviai ar saugiai klasei.

Remiantis anksčiau pasiektais agresyvaus ir saugaus vairavimo klasifikavimo rezultatais tyrimui naudojami tik 3 požymių atrinkimo/transformavimo algoritmai: 1) atsitiktinių miškų indeksai; 2) RELIEFF indeksai, bei 3) PCA požymiai. Kaip klasifikatorius pasirinktas neuroninis tinklas. Juo pasiekti rezultatai beveik nesiskyrė nuo SVM, be to kadangi duomenų taškų skaičius šiuo atveju yra ypač didelis, SVM modelio naudojimas nėra labai praktiškas.

3.2.1 pav. pateikti klasifikavimo rezultatai pridedant po vieną pagal rangą rūšiuotą požymį. Kaip matyti šiuo atveju geriausi rezultatai pasiekti naudojant atsitiktinių miškų požymius. Naudojant pirmus 7 RF požymius pasiekiamas ~77% vairuotojų tarpusavio atsiskiriamumas. Naudojant 20 rūšiuotų RF požymių pasiekiamas 79,2% klasifikavimo tikslumas, o naudojant visus požymius – 79,4%.





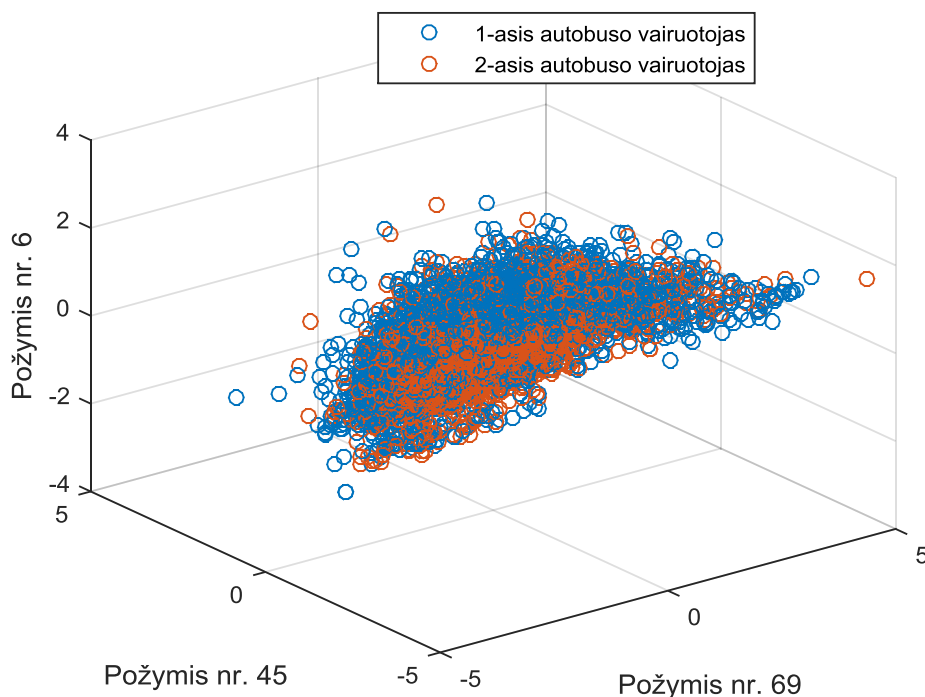
3.2.1 pav. NN autobuso vairuotojų elgsenos klasifikavimo rezultatai.

Taigi šiuo atveju požymių atrinkimas neleidžia pagerinti klasifikavimo rezultatų, bet naudojant 7 požymius pasiekiamas aukštas ~77% klasifikavimo tikslumas. Toliau išvardinami šie požymiai eilės tvarka (nuo geriausio, požymio numerio reikšmę žr. 2.5.1 lentelėje):

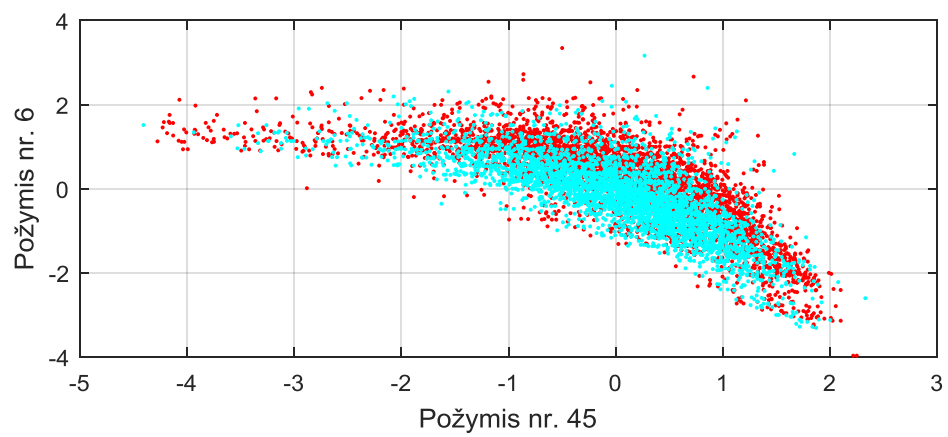
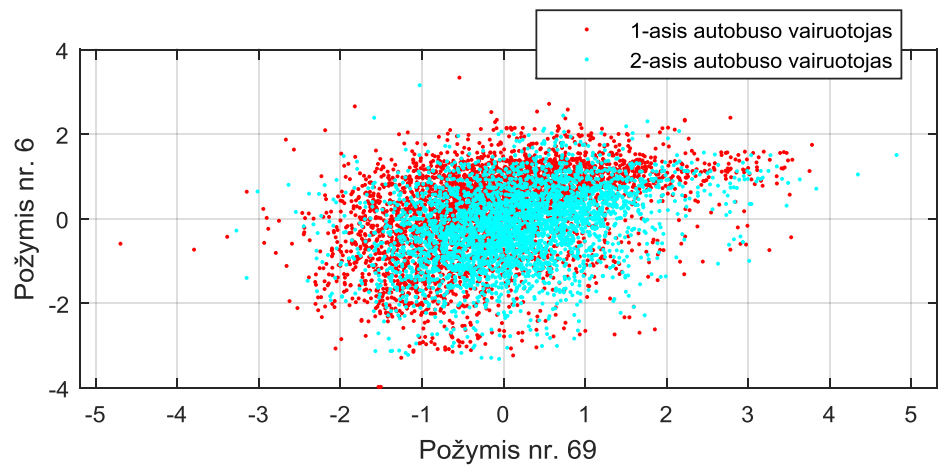
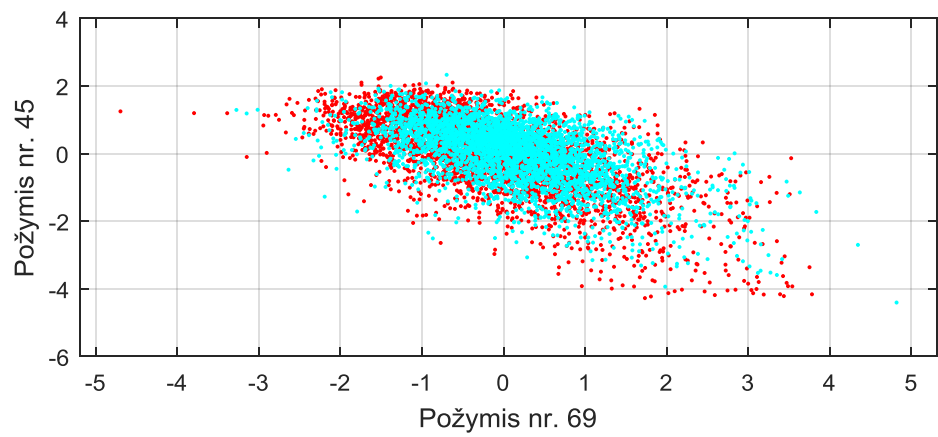
- požymis nr. 69. Šis požymis buvo atrinktas į geriausių požymių ketvertą klasifikuojant vairavimo stilių (agresyvų ir saugų vairavimą). Jis rodo spektro pirmojo dešimtadalio vidutinį intensyvumą, didesnis intensyvumas būdingas agresyvesniam vairavimui.
- požymis nr. 45. Šis požymis buvo atrinktas į geriausių požymių ketvertą klasifikuojant vairavimo stilių (agresyvų ir saugų vairavimą). Šio požymio didelės vertės rodo agresyvaus stabdymo ir akseleravimo bruožus.
- požymis nr. 6. Panašus į jį požymis nr. 3 buvo atrinktas į geriausių požymių ketvertą klasifikuojant vairavimo stilių (agresyvų ir saugų vairavimą). Tai X ašies signalo 0,2 kvartilis (požymis nr. 3 – 0,1 kvartilis). Didelės vertės parodo agresyvaus akseleravimo pobūdį.

- požymis nr. 33. Tai atstumas tarp 0,2 bei 0,8 X ašies signalo kvartilių. Analogiškas standartiniam nuokrypiui. Didelės šio požymio vertės rodo didelę dispersiją tam tikrame intervale.
- požymis nr. 3. Šis požymis buvo atrinktas į geriausių požymių ketvertą klasifikuojant vairavimo stilių (agresyvų ir saugų vairavimą). Nurodo X ašies signalo 0,1 kvartilio vertę.
- požymis nr. 48. Šis požymis buvo atrinktas į geriausių požymių ketvertą klasifikuojant vairavimo stilių (agresyvų ir saugų vairavimą). Tai trūkių verčių standartinis nuokrypis intervale. Didelės vertės rodo vairuotojo agresyvumą, keleivio diskomfortą ir didesnius kuro sąnaudas.
- požymis nr. 44. Tai Z ašies signalo duomenų ribos viršijimo intensyvumas padaugintas iš viršijančių amplitudžių vertės. Dideles šio požymio vertes galima apibūdinti kaip kelio dangos duobėtumą arba vairuotojo nenorą išvengti duobių. Šis signalas taip pat gali būti susijęs su agresyviu akseleravimu bei stabdymu dėl X ir Z ašių signalų koreliacijos akseleravimo ir stabdymo metu (priklausomai nuo amortizatorių būklės).

Vairuotojų elgsenos klasių trijų dimensijų bei porinės dviejų dimensijų vizualizacijos naudojant pirmus 3 požymius pateiktos atitinkamai 3.2.2 ir 3.2.3 paveiksluose.



3.2.2 pav. Vairuotojų elgsenos klasių pasiskirstymo vizualizacija naudojant tris geriausius požymius.



3.2.3 pav. Vairuotojų elgsenos klasių pasiskirstymo vizualizacija naudojant tris geriausius požymius (poromis).

## IŠVADOS

Atlikus išsamią literatūros analizę inercinių signalų panaudojimo intelektualiose vairavimo elgsenos identifikavimo sistemose tema, pastebėta jog dauguma sistemų neapsieina be GPS signalų vairuotojo manevrams bei stiliui identifikuoti. Dauguma egzistuojančių sistemų yra sukurtos kaip realaus laiko valdymo sistemos, suformuojančios grįžtamą ryšį (valdymo signalą) vairuotojui.

Pastebėta jog dabartiniu metu inercinių signalų panaudojimas atliekamas tiesiogiai iš išmanaus mobilaus įrenginio, kurį turi bene kiekvienas vairuotojas. Šių įtaisų panaudojimo perspektyva nekelia abejonių ir kaip tyrimo rezultatai rodo, duoda pakankamus tikslumo rezultatus jų pritaikymui įvairiuose intelektualiose transporto sistemose. Taip pat pagrindinis šių sistemų privalumas yra ne tik kaip sensorių platformos, bet ir kaip skaičiavimo ir surinkimo įrangos: skaičiavimai gali būti atliekami tame pačiame įrenginyje be papildomos įrangos, o patogus GUI leidžia sukurti vartotojo sąsajas, jau nekalbant apie duomenų persiuntimo ir priėmimo tarp tinklų galimybes.

Nors ir minėtų mobilių išmanių įrenginių pritaikymas atrodo pigus paprastas bei patogus, kai kuriais atvejais jis vis tiek išlieka brangus ir nepatogus: kada reikalinga tik jutiklių informacija ir nereikalinga vartotojo sąsaja, patogiau montuoti inercinius jutiklius prie korpuso, pvz., viešojo transporto sektoriuje, o mobilaus telefono pritaikymas gali kelti papildomų rūpesčių.

Šiame darbe buvo atlikti dvejopi tyrimai siekiant sudaryti algoritmus intelektualiai vairavimo elgsenos įvertinimo sistemai jos kūrimo metu: 1) agresyvios ir saugios vairavimo elgsenos atpažinimo; 2) vairuotojų elgsenos atpažinimo.

Agresyvios ir saugios vairavimo elgsenos atpažinimo metu nustatyta, jog naudojant 10 pagal informatyvumo rangus rūšiuotų požymių rinkinį su SVM klasifikatoriumi galima pasiekti ~95% agresyvios ir saugios vairavimo elgsenos atpažinimo tikslumą. Taip pat, naudojant vos 4 geriausius požymius galima pasiekti vidutiniškai (per visus klasifikatorius) ~91,5% atpažinimo tikslumą. Nustatyta, jog šie 4 požymiai yra susiję tik su G – jutiklio X ašimi (išilginiu pagreičiu).

Vairuotojų elgsenos atpažinimo tyrimui buvo panaudoti ilgo laikotarpio (~12 savaičių) duomenys surinkti iš viešojo transporto autobuso, kai jį įvairiais maršrutais, bei įvairiomis pamainomis vairavo du vairuotojai. Požymiai buvo išgaunami tie patys (ir iš tų pačių signalų) kaip ir agresyvaus ir saugaus vairavimo klasifikavimo atveju. Bet šiuo atveju žymės atitiko ne agresyvų ar saugų vairavimą, o vairuotoją. Nustatyta, jog vairuotojų elgsenos atpažinimas gali būti pasiekiamas ~79,4% naudojant visus požymius, ~79,2% naudojant 20 atrinktų požymių, bei

~77% naudojant vos 7 geriausias požymius. 6 iš šių 7 geriausių požymių buvo išgaunami iš X ašies signalo duomenų. Taip pat šie požymiai atitiko agresyvaus ir saugaus klasifikavimo metu atrinktus geriausias 4 požymius, kas rodo jog vairuotojų klasifikavimo uždavinys buvo panašus į agresyvaus ir saugaus vairavimo atpažinimo uždavinį. O tai savo ruožtu rodo, jog vairuotojų elgsenos skirtumai yra sąlygoti skirtingo vairavimo agresyvumo laipsnio.

## LITERATŪRA

1. GOTO H., ABE K., MUNEKATA F., KOBAYASHI K., WATANABE K. Estimation of Driving Loci and Evaluation of Driving Skill. *Proceedings of the Intelligent Vehicles '95 Symposium*, 1995, p. 388-393.
2. WANG F., MA N., INOOKA, H. A Driver Assistant System for Improvement of Passenger Ride Comfort through Modification of Driving Behaviour. *International Conference on Advanced Driver Assistance Systems (ADAS)*, 2001, p. 38-42.
3. MITROVIC D. Reliable Method for Driving Events Recognition. *IEEE Transactions on Intelligent Transportation Systems*, 2005, vol. 6, no. 2, p. 198-205.
4. MOHAN P., PADMANABHAN V. N., RAMJEE R. Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. *SenSys '08 Proceedings of the 6th ACM conference on Embedded network sensor systems*, 2008, p. 323-336.
5. ERIKSSON J., GIROD L., HULL B., NEWTON R., MADDEN S., BALAKRISHNAN H. The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. *MobiSys '08 Proceedings of the 6th international conference on Mobile systems applications and services*, 2008, p. 29-39.
6. ISHIKAWA N., ONDA K., WATANABE K., KOBAYASHI K., KURIHARA Y. Mobile phone application for ecodriving. *Proceedings of SICE Annual Conference*, 2010, p. 2424-2447.
7. DAI J., TENG J., BAI X., SHEN Z., XUAN D. Mobile Phone Based Drunk Driving Detection. *4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, 2010, p. 1-8.
8. JOHNSON D. A., TRIVEDI M. M. Driving Style Recognition Using a Smartphone as a Sensor Platform. *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011, p. 1609-1615.
9. ARAUJO R., IGREJA A., CASTRO R., ARAUJO R. E. Driving Coach: a Smartphone Application to Evaluate Driving Efficient Patterns. *IEEE Intelligent Vehicles Symposium (IV)*, 2012, p. 1005-1010.
10. EREN H., MAKINIST S., AKIN E., YILMAZ A. Estimating Driving Behavior by a Smartphone. *IEEE Intelligent Vehicles Symposium (IV)*, 2012, p. 234-239.
11. SONG C., WU J., LIU M., GONG H., GOU B. RESen: Sensing and Evaluating the Riding Experience based on Crowdsourcing by Smart Phones. *Eighth International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, 2012, p. 147-152.

12. FAZEEN M., GOZICK B., DANTU R., BHUKHIYA M., GONZALEZ M. C. Safe Driving Using Mobile Phones. *IEEE Transactions on Intelligent Transportation Systems*. 2012, vol. 13, no. 3, p. 1462-1468.
13. CASTIGNANI G., FRANK R., ENGEL T. Driver Behavior Profiling using Smartphones. *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, p. 552-557.
14. ALMAZAN J., BERGASA L. M., YEBES J. J., BAREA R., ARROYO R. Full auto-calibration of a smartphone on board a vehicle using IMU and GPS embedded sensors. *IEEE Intelligent Vehicles Symposium (IV)*, 2013, p. 1374-1380.
15. CHAKRAVARTY T., GHOSE A., BHAUMIK A., CHOWDHURY A. MobiDriveScore – A System for Mobile Sensor Based Driving Analysis. *Seventh International Conference on Sensing Technology (ICST)*, 2013, p. 338-344.
16. GEORGE T.K., GADHIA H. M., SUKUMAR R.S. O., CABIBIHAN J.-J. Sensing discomfort of standing passengers in public rail transportation systems using a smart phone. *10th IEEE International Conference on Control and Automation (ICCA)*, 2013, p. 1509-1513.
17. PFRIEM M., GAUTERIN F. Employing Smartphones as a Low-Cost Multi Sensor Platform in a Field Operational Test with Electric Vehicles. *47th Annual Hawaii International Conference on System Science (HICSS)*, 2014, p. 1143-1152.
18. AMATA H., MIYAJIMA C., OZAKI A., NISHINO T., KITAOKA N., TAKEDA K. Abrupt Steering Detection Based on the Road Construction Ordinance and Vehicle Acceleration Captured with Drive Recorders. *3rd International Conference on Innovative Computing Information and Control (ICICIC '08)*, 2008, p. 15-18.
19. SATHYANARAYANA A., BOYRAZ P., HANSEN J. H. L. Driver Behavior Analysis and Route Recognition by Hidden Markov Models. *IEEE International Conference on Vehicular Electronics and Safety (ICVES 2008)*, 2008, p. 276-281.
20. DI LECCE V., CALABRESE M. Experimental System to Support Real-Time Driving Pattern Recognition. *4th International Conference on Intelligent Computing (ICIC 2008)*, 2008, p. 1192-1199.
21. BOONMEE S., TANGAMCHIT P. Portable Reckless Driving Detection System. 6th International Conference on Electrical Engineering/Electronics. *Computer, Telecommunications and Information Technology (ECTI-CON 2009)*, 2009, vol. 1, p. 412-415.
22. LU M., WANG J., LI K., YAMAMURA T., KUGE N., NAKAGAWA T. Classification of Longitudinal Driving Behaviour Based on Simulator Study. *Proceedings of 22nd ICTCT workshop*, 2009, p. 1-11.

23. NAITO A., MIYAJIMA C., NISHINO T., KITAOKA N., TAKEDA, K. Driver evaluation based on classification of rapid decelerating patterns. *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, 2009, p. 108-112.
24. MIYAJIMA C., UKAI H., NAITO A., AMATA H., KITAOKA N., TAKEDA K. Driver risk evaluation based on acceleration, deceleration, and steering behavior. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, p. 1829-1832.
25. ALJAAFREH A., ALSHABATAT N., NAJIM AL-DIN M. S. Driving Style Recognition Using Fuzzy Logic. *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, 2012, p. 460-463.
26. SATHYANARAYANA A., SADJADI S. O., HANSEN J. H. L. Leveraging Sensor Information from Portable Devices towards Automatic Driving Maneuver Recognition. *15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2012, p. 660-665.
27. LI K., LU M., LU F., LV Q., SHANG L., MAKSIMOVIC D. Personalized Driving Behavior Monitoring and Analysis for Emerging Hybrid Vehicles. *Proceedings of the 10th international conference on Pervasive Computing (Pervasive'12)*, 2012, p. 1-19.
28. KROTAK T., SIMLOVA M. The Analysis of the Acceleration of the Vehicle for Assessing the Condition of the Driver. *IEEE Intelligent Vehicles Symposium (IV)*, 2012, p. 571-576.
29. HIGGS B., ABBAS M. A. two-step segmentation algorithm for behavioral clustering of naturalistic driving styles. *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, p. 857-862.
30. VAN LY M., MARTIN S., TRIVEDI M. M. Driver Classification and Driving Style Recognition using Inertial Sensors. *IEEE Intelligent Vehicles Symposium (IV)*, 2013, p. 1040-1045.



## **PRIEDAI**

## **Driving Style Analysis using Spectral Features of Accelerometer Signals**

**G. Zylis, V. Vaitkus, P. Lengvenis**

*Faculty of Electrical and Electronics Engineering, Kaunas University of Technology, Lithuania, E-mail: gediminas.zylis@ktu.edu*

### **Abstract**

The paper presents analysis of 3-axis accelerometer driving data in order to estimate driving style. We assume that driving style could be roughly divided into two groups: aggressive driving and safe driving. Methodology used in this research is applicable to shuttle transport driving analysis and gives possibility to analyze driving style of different road segments. The results of this work show that longitudinal acceleration data could be used in order to effectively classify aggressive and safe driving trip by using short-time Fourier transform features of different road segments.

**KEYWORDS:** *driving style, 3-axis accelerometer, driver classification, spectral analysis.*

### **1. Introduction**

The analysis of driving style for the companies, such as public transport, delivery service or insurance could help to manage service quality and increase yield by giving the possibility to control or monitor driver behavior which is related to appropriate driving style.

Signals related to driving behavior could be divided into 3 main groups: vehicle's signals, driver's signals and environment signals. Vehicle signals are obtained from dynamics of vehicle, states and outputs (w.r.t. vehicle). Driver's signals are obtained from driver's state (physiological signals) and output (driver's operating signals). Environment signals are signals that could be obtained from the traffic environment, such as states of surrounding vehicles, road signals (line markings, terrain, potholes, condition).

There are numerous researches done related to driving behavior analysis. In paper [1, 2] the experiment with instrumented vehicle was done with different drivers and analysis of driver's operating signals (pressure of accelerator and brake pedals of the vehicle) by modeling each individual driver using Gaussian Mixture Models (GMM) showed that biometric driver identification could be done using those signals' dynamic features combined with state and that they carry individual driver biometric information the most comparing other signals. Further researches of the same group of scientist [3] show that increase in driver identification rate is obtained using spectral features (i.e., cepstrum) of brake and accelerator pedal pressure signals.

In [4] the experiment with driving behavior was done using in-vehicle CAN-bus signals. The research results showed that using available CAN-bus signals (such as steering wheel angle, brake/acceleration status and vehicle's speed) using Hidden Markov Models (HMM) combined with GMM identification of driving maneuvers could be obtained without other external sensory data. Much later, using same vehicle, the experiment was done [5] that compared available CAN-bus signals and smart portable off-the-shelf device (i.e., tablet PC) sensory information. The results show that using portable device as a sensor platform the driving maneuver identification rate was higher than using CAN-bus information.

The use of smart portable device's sensors for Driver Assistance Systems (DAS) emerges rapidly because of relatively rich sensor information available and that no additional external computational hardware is necessary. It is also affected by the fact that more and more people are using smartphones and DAS could be implemented without additional hardware as an application program only which is very cheap and easy to apply. For example, in the smartphone as a sensor platform was applied in the [6] for aggressive driving and driving event recognition. The research of vehicle's condition evaluation, road condition classification and driving behavior evaluation system also using smartphone was done in [7]. Both experiment used smartphone inertial measurement sensors (accelerometer or gyroscope) and GPS.

In this work we consider that driving style could be divided into two main groups: aggressive and safe driving (regardless the land vehicle type). Aggressive driving should be considered driving type containing all negative driving features that contribute to reckless driving, vehicle damaging, passengers' discomfort and also fuel consumption. Safe driving should be considered driving type opposite to aggressive driving. Motivated by the previous experiments using vehicle's inertial signals [6-10] in this experiment that show the effectiveness of inertial sensors in driving style capturing, we use only 3-axis accelerometer data to classify driving style into aggressive and safe driving.

Our experiment constraints are as follows: the same driver is driving the same vehicle in aggressive and safe driving style on the same route. The research results application will be related to shuttle transport services when the vehicle is driven the same route all the time and no different route consideration is necessary. Spectral analysis is used in order to obtain features from accelerometer signals that later are used for driving style classification into aggressive and safe.

Further the paper is divided into following sections: 1) Accelerometer signal pre-processing; 2) Spectral features extraction; 3) Feature selection; 4) Driving style classification; 5) Conclusions and future work.

## 2. Accelerometer signal pre-processing

In this section a signal pre-processing is described, that is necessary for appropriate further use of the signal. In Fig. 1 two examples of safe/normal and aggressive driving styles are depicted of all accelerometer (G-sensor) signals.

In depicted example (Fig. 1), signal values at the end (from 16,000<sup>th</sup> to 20,000<sup>th</sup> safe/normal driving signal discrete values and from 12,500<sup>th</sup> to 15,000<sup>th</sup> aggressive driving signal discrete values) correspond to inactive driving period which doesn't contain information about driving style. This kind of signal end information should be removed. The simple thresholding methodology cannot be used for this kind of filtering because of possible data spikes (Fig. 1 aggressive driving discretely >14,000) and bias values of signal in stop period (as depicted in Fig. 2). The bias values occur because the vehicle can stop in any position w.r.t. gravity vector and because the accelerometer (G-sensor) values are measured w.r.t. gravity vector, the bias values depend on slope of the platform the vehicle is stopped (standstill on the uphill or downhill).

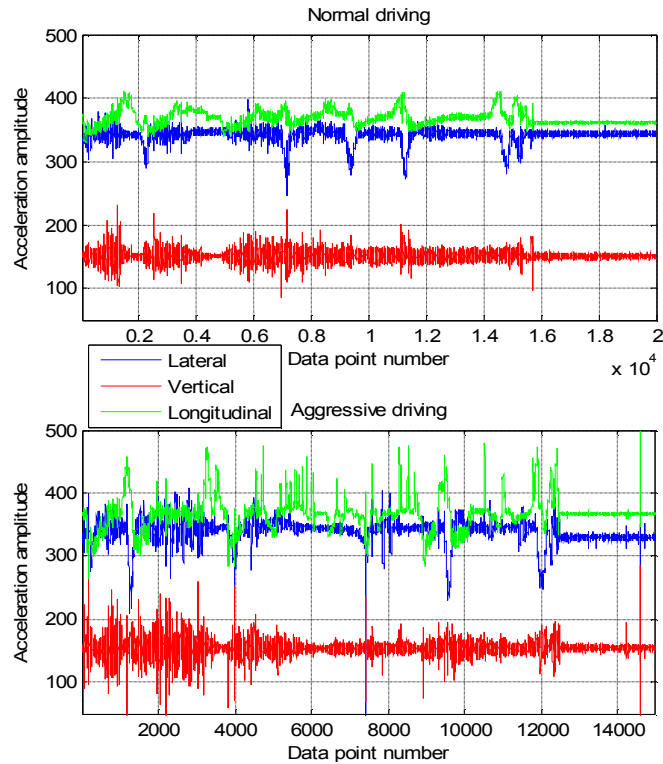


Fig. 1. 3-axis accelerometer signal examples of driving in aggressive and safe driving styles the same route. Signal values at the end correspond to inactive period (when vehicle is not moving) and must be discarded for further analysis

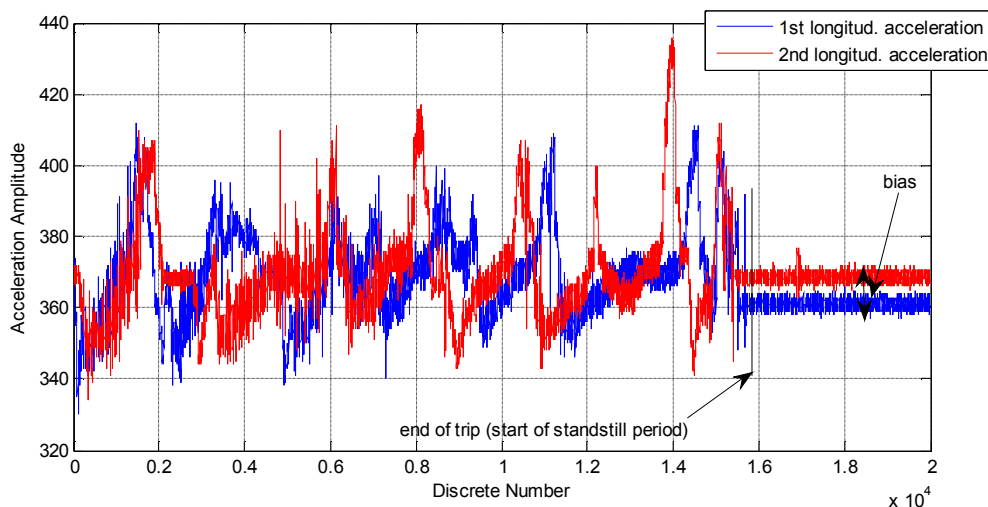


Fig. 2. Longitudinal acceleration signals of two trips of the same route. Bias is depicted as difference of two signals when vehicle standstill

In order to solve the pre-processing problem mentioned above we use median sliding window filter methodology for filtering three accelerometer signals at once. The signal end sections after proposed filtering are depicted in Fig. 3. The automatic median filtering technique removes signal values using sliding window until the minimum median difference threshold value of signal window is violated (minimum median difference value is 1). The tradeoff between window length and precise median estimation exist, in this experiment we use 100 discrete sliding window.

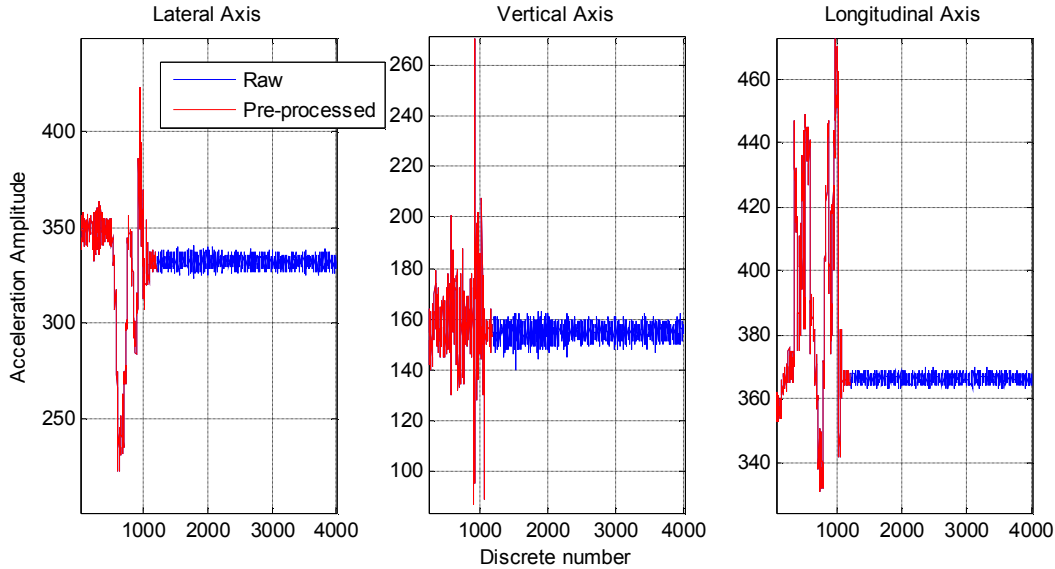


Fig. 3. Raw acceleration signals and filtered signals using median sliding window methodology

### 3. Spectral feature extraction

The aggressiveness of driving style is related to variability of the acceleration signals, it could be noticed in Fig. 1 example that when driving aggressively the range of signal and variability increases. From the first look, when analyzing signal in time domain, it could be reasonable to determine some threshold values and calculate the times the signal violate threshold values. But, as mentioned before, the acceleration signal amplitude are calculated w.r.t. gravity vector and if vehicle is going uphill for example, the bias value occur and simple thresholding doesn't work. Therefore we use spectral analysis: the signal variability and power (high amplitude deviations) could be captured using short-time Fourier transform. The whole signal Fourier transform can only capture the whole signal variability and no information about possible occasional aggressive driving. In order to capture aggressiveness of driving in the signal zones of interest and increase their contribution, a windowed (short-time) Fourier transform should be applied: short-time significant aggressiveness of signal could be obtained and so more precisely the driving style could be estimated taking into consideration short-time signal intervals of the whole trip. Before that, signal resampling were performed to have same amount of data for each signal. The short-time Fourier transform (in continuous time domain) is defined as follows:

$$F(\omega, \tau) = \int f(t)w(t-\tau)e^{-j\omega t} dt \quad (1)$$

where  $\omega$  is frequency;  $\tau$  is time;  $w(t-\tau)$  is window function (in our experiment we use Hamming window). After the short-time Fourier transform (using fast Fourier transform in MATLAB environment), we calculate power spectral density. The power spectral density matrix is defined as:

$$P(i, j) = k|F(i, j)|^2 \quad (2)$$

where  $F(i, j)$  is discrete-time fast Fourier transform and  $k$  for one-sided power spectral density is defined as follows:

$$k = \frac{2}{\sum_{n=1}^L |w(n)|^2} \quad (3)$$

where  $L$  is length of window. At zero and Nyquist frequencies, the factor of two in the numerator is replaced by 1. After the power spectral density calculation the example of spectrogram (logarithm of power spectral density in time windows of short-time Fourier transform) is depicted in Fig. 4.

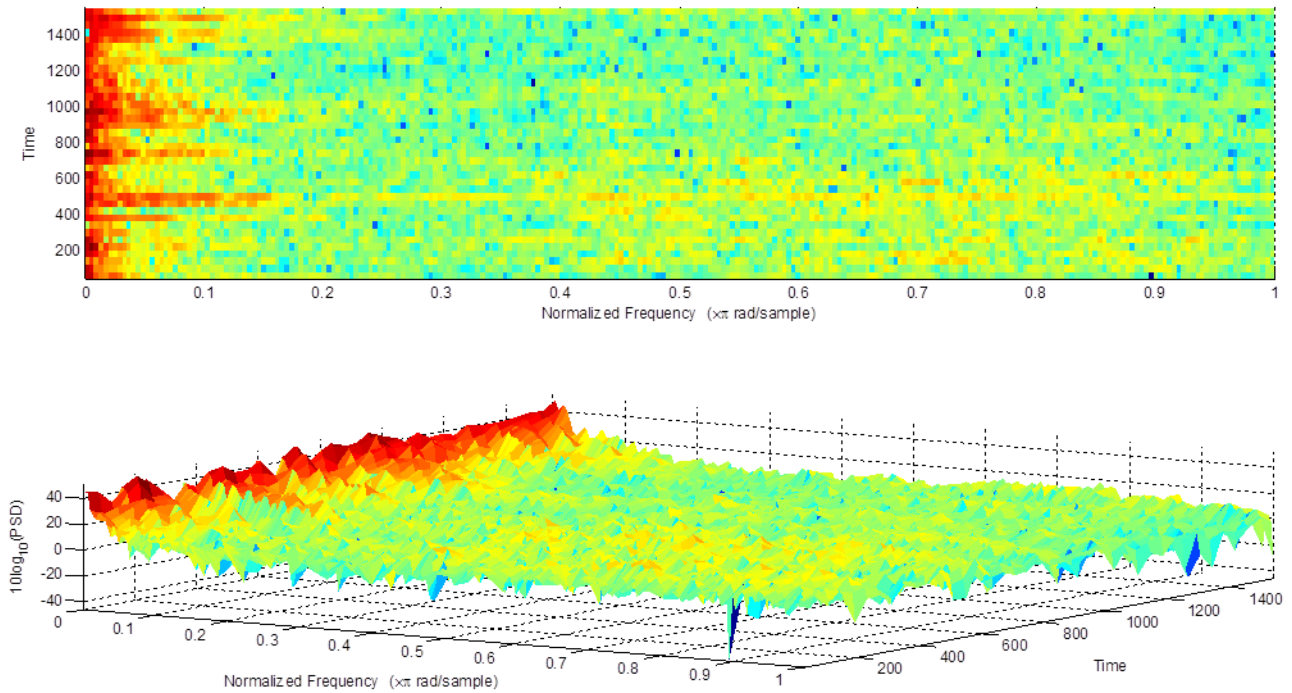


Fig. 4. An example of one axis accelerometer signal spectrogram (PSD – power spectral density) using 500 signal discrete window length with 50% overlap factor

Further we use logarithm of power spectral density of appropriate frequency values as features for each window segment and later use for classification of aggressive and safe driving segments for each driving trip.

#### 4. Feature selection

In this section, we discuss feature selection algorithm which is useful for dimensionality reduction and selected features are later used for classification. In this research we use Principal Component Analysis (PCA) methodology for dimensionality reduction. Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The example of PCA operation is given in Fig. 5.

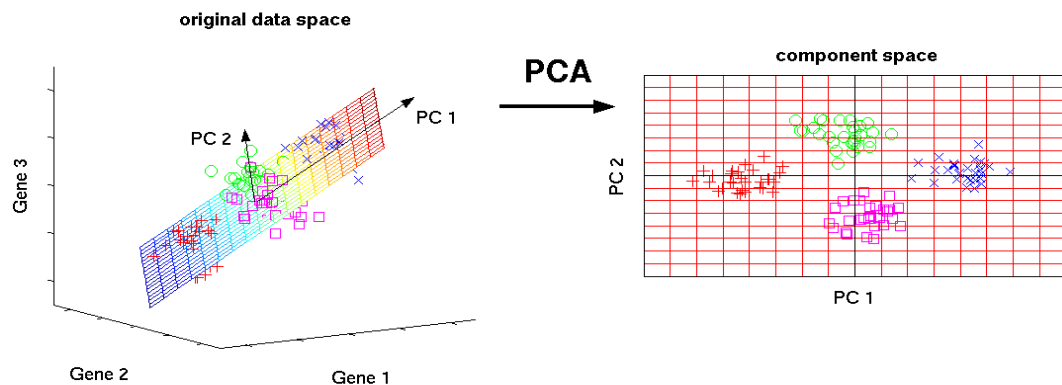


Fig. 5. An example of principal component analysis for dimensionality reduction. PCA assumes maximum variance criterion when choosing most discriminative principal components for classification task

The PCA analysis using various combinations of signal features are graphically illustrated in Fig.6. Depicted segments were selected using reference aggressive driving and safe driving signals. From Fig. 6 the discriminative capabilities can be seen for various combinations selecting first two principal components for visualization. It is clear that longitudinal acceleration data alone can reasonably good be used for classification, however still combination of longitudinal and lateral acceleration, longitudinal and vertical acceleration and all three signal combination could be also reasonable choice. Other combinations seem to give less important results to further we investigate previous mentioned combinations.

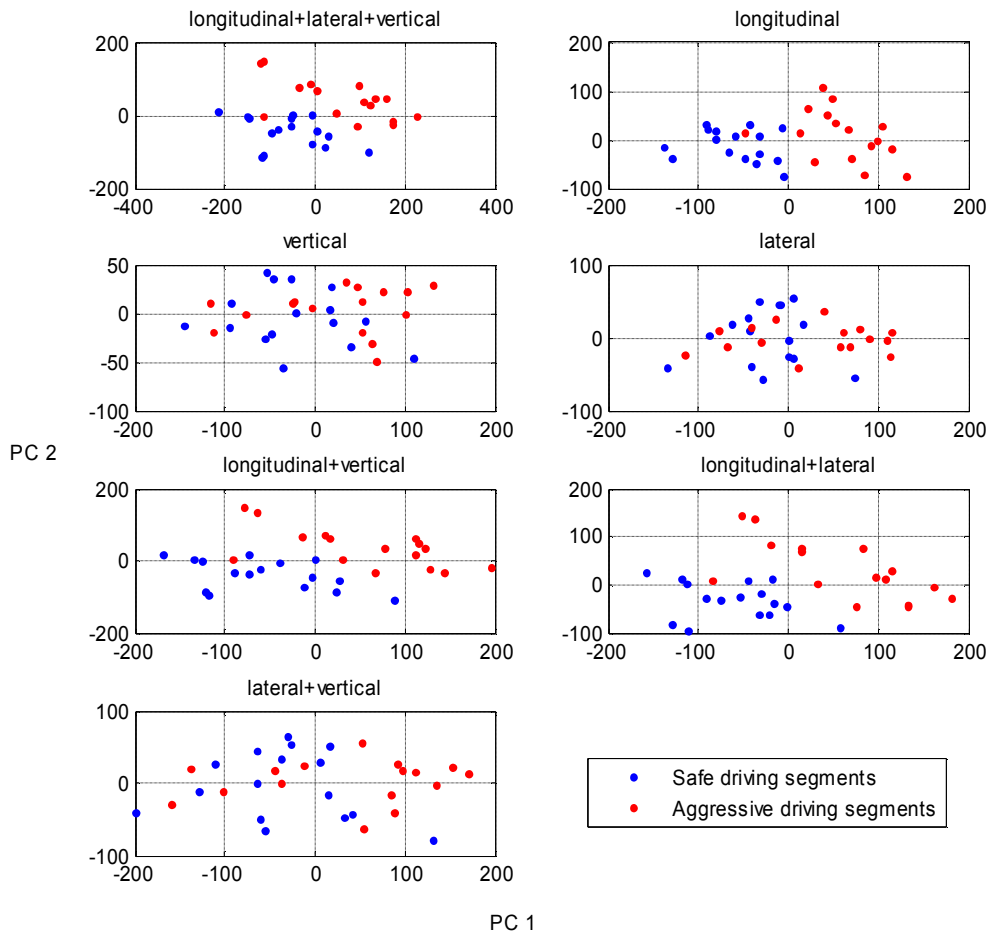


Fig. 6. Reference aggressive and safe driving segment PCA extracted features (two Principal Components (PC)) illustration after combining various signals' original features (spectrograms)

## 5. Driving style classification

In previous section we used PCA analysis to investigate what combinations of features to use further for classification. In this section we use whole driving signal (aggressive and safe) segments using moving window with 50% overlap factor in order to generate spectrogram features and classify them, when classification method is trained with previously (section 4) depicted aggressive and safe driving segments. The purpose of this classification is as follows: each separate driving trip segments' (after performing spectrogram feature generation) principal components (principal component eigenvectors are obtained from reference driving trips) are used as an input to the trained classifier in order to obtain score of driving trip (aggressiveness or safety of driving trip after classifying each driving segment into aggressive or safe). And we compare classification-based scores with the expert decisions about safe or aggressive driving trip.

For this task, we used Random Forest (RF) classifier with two principal components. No additional cross-validation checks were performed because of RF classifier (bagged decision trees in MATLAB) internal use of bagging (bootstrap aggregating) ensemble methodology. Classification results after out-of-bag error analysis favor longitudinal acceleration signal alone compared to other combinations (Fig. 7) achieving only  $\sim 3\%$  error for aggressive and safe reference segments. So further we use only longitudinal acceleration signal in order to obtain aggressiveness and safety score for each driving trip.

The results of aggressiveness and safety of driving using experimental driving trips are summarized in Tab. 1.

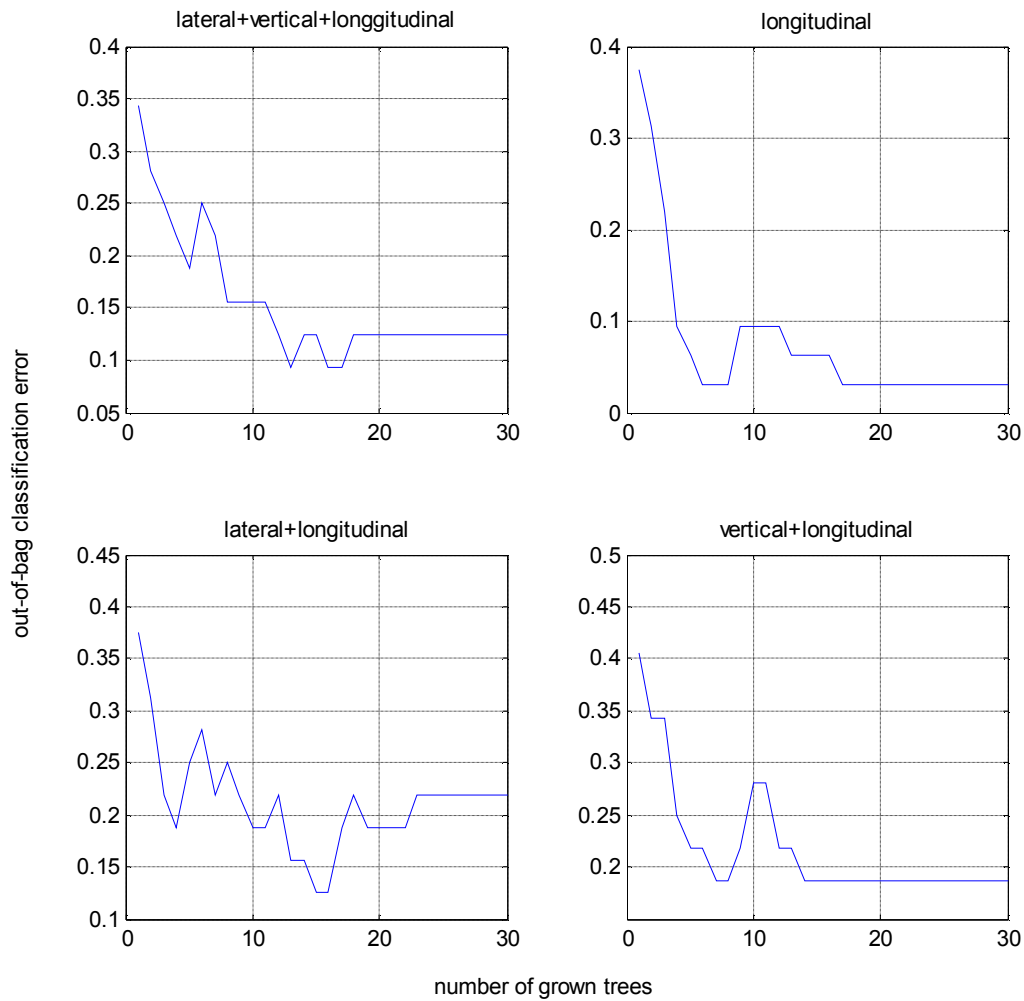


Fig. 7. Out-of-bag random forest classification error dependency on number of grown trees. The smallest misclassification percentage is using only longitudinal signal features

Table 1

Aggressive and safe driving trips classification and expert label comparison

Driving Number	Safe segments classified	Aggressive segments classified	Aggressive/safe	Expert label
1	13	26	2	aggressive
2	9	30	3.333333	aggressive
3	9	30	3.333333	aggressive
4	7	32	4.571429	aggressive
5	11	28	2.545455	aggressive
6	25	14	0.56	safe
7	28	11	0.392857	safe
8	39	0	0	safe
9	38	1	0.026316	safe
10	39	0	0	safe

## 6. Conclusions

After the comparison in Table 1, it is observed that high correlation between expert labels and aggressive/safe ratio obtained by classifying each road segment, exist: the bigger the ratio, the more aggressive driving is.

Classification methodology presented in this work could be useful for automatic aggressive driving detection and recognition. After all analysis we can conclude that for our application purposes longitudinal acceleration signal information is sufficient.

This research also shows that with appropriate constraints, driving style classification and recognition (considering aggressive and safe driving) can be done by a very low cost sensory information – longitudinal accelerometer alone and lets minimize system costs to minimum. Only requirements for the sensor is appropriate

attachment to vehicle body. This kind of system would not interfere with any vehicle system devices and is easily to integrate with other system elements.

In future, we will use more inertial signals and are planning to expand the possibilities for this kind of system by minimizing constraints and integrating more features that would increase capabilities of transport management.

## References

1. **Igarashi K., Miyajima C., Itou K., Takeda K., Itakura F. and Abut H.** Biometric Identification Using Driving Behavioral Signls. IEEE Int. Conf. Signals, Multimedia and Expo, vol. 1, Jun. 2004, p.65-68.
2. **Wakita T., Ozawa K., Miyajima C., Igarashi K., Itou K., Takeda K. and Itakura F.** Driver Identification Using Driving Behavior Signals. IEEE Proc. Intelligent Transportation Systems, Sept. 2005, p.396-401.
3. **Miyajima C., Nishiwaki Y., Ozawa K., Wakita T., Itou K. and Takeda K.** Cepstral Analysis of Driving Behavioral Signals for Driver Identification. IEEE Int. Conf. Acoustics, Speech and Signal Processing, vol. 5, May 2006, p.921-924.
4. **Choi S., Kim J., Kwak D., Angkititrakul P. and Hansen J. H. L.** Analysis and Classification of Driver Behavior using In-Vehicle CAN-Bus Information. 2007.
5. **Sathyanarayana A., Sadjadi S. O. and Hansen J. H. L.** Leveraging Sensor Information from Portable Devices towards Automatic Driving Maneuver Recognition. IEEE Int. Conf. Intelligent Transportation Systems, Sept. 2012, p. 660-665.
6. **Johnson D. A. and Trivedi M. M.** Driving Style Recognition Using a Smartphone as a Sensor Platform. IEEE Int. Conf. Intelligent Transportation Systems, Oct. 2011, p. 1609-1615.
7. **Fazeen M., Gozick B., Dantu R., Bhukhiya M., Gonzalez M. C.** Safe Driving using Mobile Phones. IEEE Trans. Intelligent Transportation Systems, vol. 13, no. 3, September 2012, p. 1462-1468.
8. **Aljaafreh A., Alshabatat N. and Al-Din M. S. N.** Driving Style Recognition Using Fuzzy Logic. IEEE Int. Conf. Vehicular Electronics and Safety, Jul. 2012, p. 460-463.
9. **Ly M. V., Martin S., Trivedi M. M.** Driver Classification and Driving Style Recognition using Inertial Sensors. Intelligent Vehicles Symposium, Gold Coast, Australia, June 23–28, 2013, p. 1040-1045.
10. **Krotac T., Simlova M.** The Analysis of the Acceleration of the Vehicle for Assessing the Condition of the Driver. Intelligent Vehicles Symposium, Alcalá de Henares, Spain, June 3–7, 2012, p. 571-576.



# Driving Style Classification using Long-Term Accelerometer Information

Vygandas Vaitkus and Paulius Lengvenis

Department of Automatic  
Kaunas University of Technology  
Kaunas, Lithuania

Gediminas Žylius

Joint Stock Company “RMD Technologies”  
Kaunas, Lithuania

**Abstract**—Driving style can be characteristically divided into normal and aggressive. Related researches show that useful information about driving style can be extracted using vehicle’s inertial measurement signals with the help of GPS. However, for public transportation the GPS sensor isn’t necessary because of repetition of the route. This assumption helps to create low-cost intelligent public transport monitoring system that is capable to classify aggressive and normal driver. In this paper, we propose pattern recognition approach to classify driving style into aggressive or normal automatically without expert evaluation and knowledge using accelerometer data when driving the same route in different driving styles. 3-axis accelerometer signal statistical features were used as classifier inputs. The results show that aggressive and normal driving style classification of 100% precision is achieved using collected data when driving the same route.

**Index Terms**—Vehicle driving, intelligent vehicles, pattern recognition, accelerometer.

## I. INTRODUCTION

There are many researches done on driving behaviour, style and skill estimation. Despite the difference between concepts, information-carrying signals collected are the same: driver, vehicle or environment signals. The results obtained from one study are closely related to another. In our work we consider aggressive driving style to be both fuel-consuming and unsafe. In contrast to aggressive driving we consider normal driving style.

In our research we used inertial measurements of the vehicle, i.e., 3-axis accelerometer. This sensor is low-cost and ubiquitous (because of various smart devices). The research studies have shown that accelerometer captures driving style, skill and driving behaviour information. T. Krotak, M. Simlova [1] in experiment show that dynamic driver behaviour information could be gathered from the longitudinal and lateral acceleration signals: the features of the signal significantly differ comparing expert and beginner drivers, also the significant difference is observed comparing concentrated versus distracted (i.e., speaking phone) driver on the same road track. M. V. Ly *et al.* [2] in experiment use driving event (turn, acceleration and braking) features for driver classification. The features are extracted from gyroscope and accelerometer signal segments of driving

events for classifying two different driving styles. Their experiment also shows a high correlation between vehicle inertial measurements and direct vehicle control signals (steering wheel angle, brake/acceleration pedal operation). C. Miyajima *et al.* [3] in their research use longitudinal, lateral acceleration and velocity signals captured in drive recorders for the driver risk analysis. They compared experimental results with the risk scores estimated by experts and showed that some correlation exists. A. Aljaafreh *et al.* [4] propose a driving performance inference system based on the signature of acceleration in two dimensions and speed. Fuzzy system is used to distinguish between aggressive and normal driving styles. V. Manzoni *et al.* [5] evaluate driving style economy considering fuel consumption using mathematical modeling and information acquired from inertial signals and speed from GPS. The authors also show that 20% of economy index (that indicates fuel consumption) difference is observed comparing several public transport drivers. This experiment approves that the potential of public transport fuel consumption optimization exist. Pattern recognition approach is taken in Y. Zhang *et al.* [6] research study. Researchers use driver’s steering behaviour, i.e., the coefficients of the discrete Fourier transform (DFT) of the steering wheel angle, to identify the driving skill level. Results show that difference between expert and beginner drivers exist in frequency domain. The experiment was done using driving simulator. M. R. Othman *et al.* [11] and Y. L. Murphey *et al.* [12] present jerk analysis (first derivative of acceleration signal) that also could give good results for classification of driving style.

The latest researches in this area show the tendency of using smartphones or tablets as a sensor platform instead of conventional vehicle CAN-bus sensor information. A. Sathyanarayana *et al.* [7] compare driving maneuver recognition results using both CAN-bus and tablet information and show that using smartphone sensor information one can obtain better results (about 15% classification accuracy increase) than CAN-bus sensor information. However, the analysis was done using mostly different signals so the results don’t take into account the quality of the signals, but more likely the quantity (off-the-shelf tablets have more sensors than conventional CAN-bus connects). D. A. Johnson, M. M. Trivedi [8] created the system called MIROAD that uses Dynamic Time Warping (DTW) and smartphone based sensor-fusion (accelerometer, gyroscope, magnetometer, GPS,

video) to detect, recognize and record aggressive driving actions. M. Fazeen *et al.* [9] use the 3-axis accelerometer and GPS of an Android-based smartphone to record and analyze various driver behaviors and external road conditions. R. Araujo *et al.* [10] in experiment use smartphone to evaluate driving-efficient patterns to form a feedback as an eco-driving advice to the driver. Authors obtain CAN-bus speed, acceleration, altitude, throttle, instant engine fuel consumption, engine rotation signals using bluetooth adapter and processes them in the smartphone.

The aim of the urban public transport companies are high-quality and cost-effective transportation because they have to compete with private transporters which are faster, more accessible and more convenient in most cases. The local urban public transport company is research initiator and customer of the product, which will classify the drivers according to the driving quality. Based on the driving quality evaluation the drivers will be ranked. Worst drivers will be send on additional training and best drivers encouraged with monthly salary rising. It is possible to buy such kind of systems but they are too expensive and not appropriately tested in practice.

After the exhaustive research it was found that most of the researches and developers had a purpose to create a real-time safety or other driving estimation systems. They try to recognize driving events, segment the signal further evaluating features given to those events or signal segments in real-time. Such systems require huge computing capacity often requires connection to vehicle electrical components and this is not acceptable to the customer, because he wants a portable system that is easy to install and are invisible to the driver. Driver must be aware that driving quality evaluation system is installed in the bus. Depending on the requirements it was decided to develop the embedded system for driving data collection based on microprocessor and accelerometer. Physical size of the system is very small, and this allows it to be installed in various locations of the bus. Data could be read at the end of the route or day from the SD memory card.

This paper only considers the public transport and the repeatability of the route. And we try to classify the driving style directly obtaining various features from the whole signal of the same route instead of its segments. This research is further divided in to signal processing, feature extraction, feature selection and classification results sections.

## II. DATA PROCESSING

Data processing was done in order to make signal suitable for feature extraction task. Figure 1 illustrates the 3-axis accelerometer example data of aggressive and normal driving. The acceleration type is shown in the legend. Because we use only accelerometer and no other sensors, first we must extract useful signal from whole original signal which is driving period between beginning and ending of the trip (when the vehicle is moving). Extracted signal is used for analysis excluding redundant data (which is generated at the beginning and after completion of the track when the vehicle is not moving). For this reason, sliding window was used and

comparison of statistical central characteristic (i.e., medians) between windows was done.

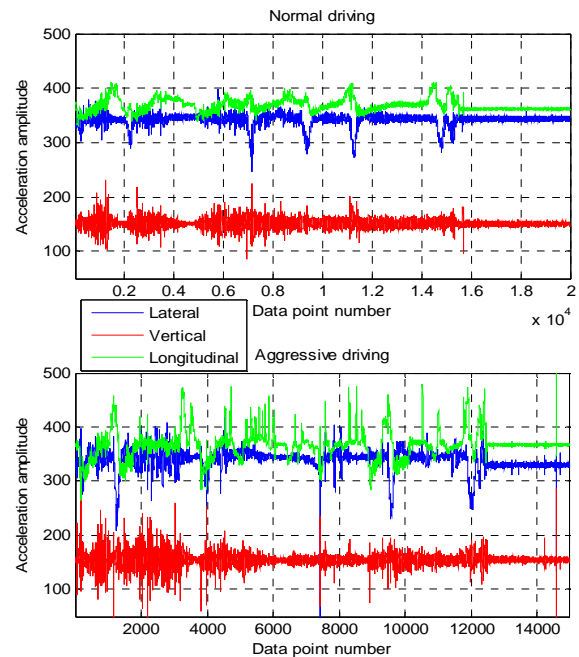


Fig. 1. Raw 3-axis accelerometer data, obtained from car when driving same track in aggressive and normal driving style.

After extraction of the useful data (signal period when the vehicle is moving), removing various high frequency components (one discrete data point width) peaks and outliers is necessary for several features that will be extracted. In this case a robust local regression technique was used with 5 data point window. In further section, we use processed data for several feature calculation and extracted data (without the redundant stop period data after track completion) for others.

## III. FEATURE EXTRACTION

The features in time domain are extracted from all three accelerometer signals (lateral, longitudinal and vertical acceleration). Various features keep various quantities of useful information about driving style. Also, it is unknown what set (quantity and type) of features most accurately separate the driving styles. So it was decided to extract almost all possible features from time domain and then use a feature selection method for selection of the most informative set.

The statistical features calculated are:

- central characteristics (mean, median, mode);
- dispersion characteristics (variance, interquartile range, mean absolute deviation, median absolute deviation, standard deviation, range);
- minimum and maximum values;
- histogram extent characteristics (skewness, kurtosis, 3<sup>rd</sup> and 4<sup>th</sup> statistical moments);

- Kendall's tau rank correlation coefficient between each pair of acceleration signals;
- covariance between each pair of acceleration signals;
- criteria of data threshold violation (DTV);
- difference between reference and ordinary driving histograms;
- polynomial parameters identified after approximation of inverse cumulative density histogram.

Above described statistical features can be calculated by using a variety of software applications, but last three of them will be described in more detail.

Data threshold violation (DTV) was calculated by selecting minimum and maximum bounds according to reference normal driving style processed signal minimum and maximum values. The threshold violation ratio was selected as a feature, calculated as the number of times  $N_1$  the data exceed bounds divided by the overall data length  $N$ , which is calculated as

$$DTV = \frac{N_1}{N}. \quad (1)$$

Figure 2 illustrates an example of boundary selection using reference normal driving compared to typical aggressive driving of longitudinal acceleration.

Figure 3 illustrates how normal and aggressive driving histograms of longitudinal acceleration differ from each other. Because the signal has only integer values so the bin widths of histograms were used as integer values of acceleration amplitude.

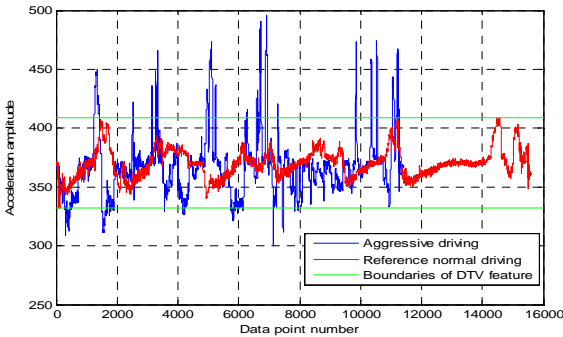


Fig. 2. Longitudinal processed data boundaries calculated from reference normal driving. Violations are observed when driving in aggressive style.

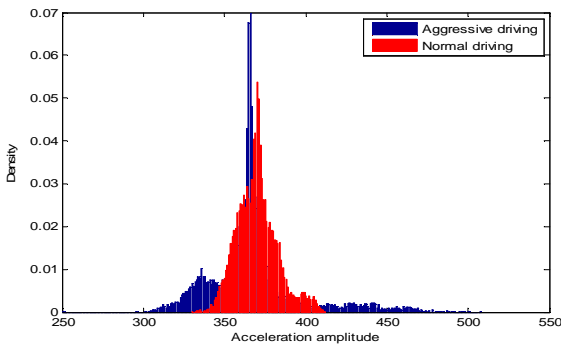


Fig. 3. Density histograms. Longitudinal acceleration density histogram comparison between aggressive and normal driving experiments.

The values of the differences between reference normal (or aggressive) histogram and experimental driving histograms bin height were calculated. Sum of calculated absolute differences were selected as one feature and the covariance between reference difference values (calculated difference values between reference normal driving and typical aggressive driving) and experimental difference values was used as another feature. The differences were calculated when histogram density values were centred according to maximum covariance values also when no centring is applied (differences calculated directly like in Fig. 3). The cross-covariance algorithm was used and the histogram density value arrays (assuming no relation between acceleration amplitudes) were centred according to maximum covariance between them.

Furthermore, we approximate inverse cumulative density histogram by polynomials and use approximating polynomial parameters as other features. We believe that those parameters  $p_{n+1-i}$  contain discriminative information for driving classification. Approximating  $n^{\text{th}}$  order polynomial model is

$$y = \sum_{i=0}^{n+1} p_{n+1-i} x^{n+1-i}. \quad (2)$$

3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup> degree polynomial parameters were used as features. The approximation was done using least-squares method.

#### IV. FEATURE SELECTION

In research 39 features are calculated for each of three acceleration signals, which conclude 117 features over all. To create all possible sets of features and evaluate their significance is very expensive in relation to computing costs, so in order to find the most discriminative features Student's t-test was performed. This test allows sorting all features according to their discriminative capability. After sorting a classification is performed using sorted features adding one by one to feature set and calculating misclassification error.

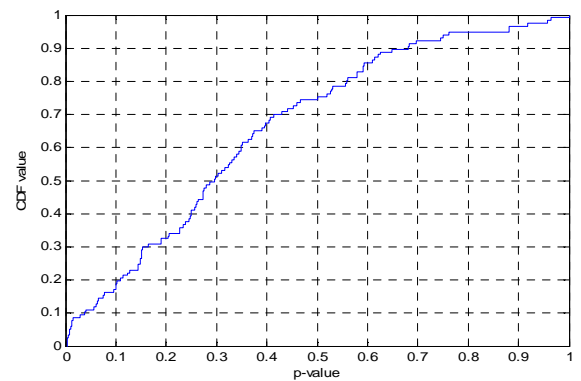


Fig. 5.  $p$ -value cumulative density function. When 5% hypothesis' significance level is chosen, there are more than 10% of all features that have high discriminative capabilities.

Figure 5 shows the discriminative capability in p-values, the lower the p-value the more discriminative the feature is. We use 0.05 p-value significance level, which means that features that have p-value less than 0.05 are considered having high discriminative capabilities. Experimental results (Figure 5) showed that more than 10% of all features have high discriminative capability. Those features were selected and sequential forward feature selection (SFFS) was performed in order to obtain subset of most discriminative features. The correlation of features problem assumed to be not important because SFFS with cross-validation is performed and even if the subset of correlated features improves classification, this subset is selected.

## V. CLASSIFICATION RESULTS

For driving style classification  $k$ -nearest neighbours classification algorithm was used. This algorithm classifies test sample according to majority votes of  $k$ -nearest neighbour, i.e., the sample is classified to that class which has the majority votes.

For experimental investigation 110 route driving samples were provided by urban public transport system. An experiment labels were defined by an expert investigation of the route signals. For testing it was used cross-validation to evaluate classification accuracy and average over the classification results.

Experimental investigation showed that maximum (100%) classification precision is achieved using following features:

- Mean of longitudinal acceleration;
- Mean of vertical acceleration;
- Median of vertical acceleration;
- Covariance between longitudinal and lateral acceleration;
- Three 5th order polynomial coefficients after inverse cumulative density histogram approximation.

## VI. CONCLUSIONS

Accuracy of 100% is achieved using only 7 of all 117 features. The interesting observations are that 98% accuracy is achieved using only central characteristics (mean and median) of the longitudinal and vertical acceleration.

Need to be mentioned, that experimental driving samples were obtained in the same season, traffic conditions and on the same route, this might cause that driving samples were too easy separable. However for our application purpose the results show that feature extraction of the whole signal on the same track instead of dynamic feature extraction of the signal segments could be used to classify aggressive and safe driving with high accuracy.

This type of analysis of the whole track signal is applicable to public transport, because routes consist of several or more tracks that are repetitive.

In the future works the track automatic segmentation will be used by using more signals to decompose various public

transport routes of several tracks and analyse driving style. This type of system could be route-independent (because a route consists of small parts from station to station and these small parts repeat in several routes). The system also should take into account various road conditions such as traffics, seasons, working hours, celebration days, weekends, vehicle parameters etc. Also, fuel economy index calculation will be investigated in order to classify drivers according fuel consumption.

This kind of system could be used not only for driving style analysis, but for sophisticated public transport monitoring system, which could evaluate road conditions in various time periods and seasons that could be useful for traffic forecasting, also for various analysis that could increase comfort and fuel economy of public transport.

## REFERENCES

- [1] T. Krotac, M. Simlova, "The Analysis of the Acceleration of the Vehicle for Assessing the Condition of the Driver", *Intelligent Vehicles Symposium*, Alcalá de Henares, Spain, June 3–7, 2012, pp. 571–576.
- [2] M. V. Ly, S. Martin, M. M. Trivedi, "Driver Classification and Driving Style Recognition using Inertial Sensors", *Intelligent Vehicles Symposium*, Gold Coast, Australia, June 23–28, 2013, pp. 1040–1045.
- [3] C. Miyajima, H. Ukai, A. Naito, H. Amata, N. Kitaoka, K. Takeda, "Driver Risk Evaluation based on Acceleration, Deceleration, and Steering Behavior", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, May 22–27, 2011, pp. 1829–1832.
- [4] A. Aljaafreh, N. Alshabat, M. S. N. Al-Din, "Driving Style Recognition using Fuzzy Logic", *IEEE International Conference on Vehicular Electronics and Safety*, Istanbul, Turkey, July 24–27, 2012, pp. 460–463.
- [5] V. Manzoni, A. Corti, P. D. Luca, S. M. Savaresi, "Driving Style Estimation via Inertial Measurements", *13<sup>th</sup> International IEEE Annual Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, September 19–22, 2010, pp. 777–782.
- [6] Y. Zhang, W. C. Lin, Y.-K. S. Chin, "A Pattern-Recognition Approach for Driving Skill Characterization", *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 905–916, December, 2012.
- [7] A. Sathyanarayana, S. O. Sadjadi, J. H. L. Hansen, "Leveraging Sensor Information from Portable Devices towards Automatic Driving Maneuver Recognition", *15<sup>th</sup> International IEEE Conference on Intelligent Transportation System*, Anchorage, Alaska, USA, September 16–19, 2012, pp. 660–665.
- [8] D. A. Johnson, M. M. Trivedi, "Driving Style Recognition using a Smartphone as a Sensor Platform", *14<sup>th</sup> International IEEE Conference on Intelligent Transportation Systems*, Washington, DC, USA, October 5–7, 2011, pp. 1609–1615.
- [9] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, M. C. Gonzalez, "Safe Driving using Mobile Phones", *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1462–1468, September 2012.
- [10] R. Araujo, A. Igreja, R. de Castro, R. E. Araujo, "Driving Coach: a Smartphone Application to Evaluate Driving Efficient Patterns", *Intelligent Vehicles Symposium*, Alcalá de Henares, Spain, June 3–7, 2012, pp. 1005–1010.
- [11] M. R. Othman, Z. Zhang, T. Imamura, T. Miyake, "A Study of Analysis Method for Driver Features Extraction", *IEEE International Conference on Systems, Man and Cybernetics*, October 12–15, 2008, pp. 1501–1505.
- [12] Y. L. Murphey, R. Milton, L. Kiliaris, "Driver's Style Classification using Jerk Analysis", *IEEE Symposium on Computational Intelligence in Vehicles and Vehicular Systems*, Nashville, TN, USA, March 30–April 2, 2009, pp. 23–28.