

Model rane procjene kvarova rotacijske opreme primjenom dubokog strojnog učenja

Kolar, Davor

Doctoral thesis / Disertacija

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:235:773286>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-12-24**

Repository / Repozitorij:

[Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb](#)





Sveučilište u Zagrebu

FAKULTET STROJARSTVA I BRODOGRADNJE

Davor Kolar

**MODEL RANE PROCJENE KVAROVA
ROTACIJSKE OPREME PRIMJENOM
DUBOKOG STROJNOG UČENJA**

DOKTORSKI RAD

Zagreb, 2019.



Sveučilište u Zagrebu

FAKULTET STROJARSTVA I BRODOGRADNJE

Davor Kolar

**DEEP LEARNING-BASED EARLY FAULT
DIAGNOSIS MODEL FOR ROTARY
MACHINERY**

DOCTORAL DISERTATION

Zagreb, 2019.



Sveučilište u Zagrebu

FAKULTET STROJARSTVA I BRODOGRADNJE

Davor Kolar

**MODEL RANE PROCJENE KVAROVA
ROTACIJSKE OPREME PRIMJENOM
DUBOKOG STROJNOG UČENJA**

DOKTORSKI RAD

Mentor:

Prof.dr.sc. Dragutin Lisjak

Zagreb, 2019.



Sveučilište u Zagrebu

FAKULTET STROJARSTVA I BRODOGRADNJE

Davor Kolar

**DEEP LEARNING-BASED EARLY FAULT
DIAGNOSIS MODEL FOR ROTARY
MACHINERY**

DOCTORAL DISERTATION

Supervisor:

Prof. Dragutin Lisjak, PhD

Zagreb, 2019.

Podaci za bibliografsku karticu

UDK	
Ključne riječi	
Znanstveno područje	Tehničke znanosti
Znanstveno polje	Strojarstvo
Institucija	Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje
Mentor	Prof.dr.sc. Dragutin Lisjak
Broj stranica	173
Broj slika	51
Broj tablica	27
Broj korištenih bibliografskih izvora	103
Datum obrane	
	Izv.prof.dr.sc. Danko Brezak
	Prof.dr.sc. Dragutin Lisjak
	Prof.dr.sc. Danijel Pavković
Povjerenstvo	Doc.dr.sc. Tomislav Staroveški
	Prof.dr.sc. Tomislav Šarić (<i>Sveučilište J. J. Strossmayer u Osijeku, Strojarski fakultet u Slavonskom Brodu</i>)
	Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje
Institucija u kojoj je rad pohranjen	Nacionalna i sveučilišna knjižnica u Zagrebu

Zahvala

Zahvaljujem mentoru, prof. dr.sc. Dragutinu Lisjaku, na savjetima, razumijevanju, bezrezervnoj podršci te na svim sugestijama od početne ideje do završetka provedbe istraživanja i pisanja ove disertacije.

Zahvaljujem članovima *Povjerenstva* na izdvojenom vremenu za konzultacije i konstruktivnim savjetima.

Zahvalio bih se djelatnicama i djelatnicima *Fakulteta strojarstva i brodogradnje* te svog *Zavoda za industrijsko inženjerstvo* na kolegijalnosti i podršci tijekom cijelog poslijediplomskog studija.

Riječi zahvale upućujem i kolegi Mihaelu Gudlinu, mag.ing.mech., na nesebičnoj pomoći pri shvaćanju matematičke pozadine potrebne za razumijevanje izrade modela.

Veliko hvala mom prijatelju, kolegi i kumu dr.sc. Mihi Klaiću na savjetima, motiviranju kao i pomoći pri savladavanju administrativnih zadataka.

Na kraju, najveću zahvalnost dugujem cijeloj svojoj obitelji, pri čemu posebnu zahvalu na podršci i razumijevanju upućujem voljenoj supruzi Andreji.

Sažetak

Trend razvoja industrije prema konceptu 4.0 u području održavanja promovira ideju donošenja odluka o aktivnostima održavanja temeljenih na analitici velike količine podataka.

Tradicionalni algoritmi strojnog učenja omogućavaju opisivanje i otkrivanje složenih odnosa unutar velike količine podataka, no pritom zahtijevaju ekspertno znanje iz domene problema tijekom obrade i analize ulaznih signala s ciljem određivanja relevantnih značajki na temelju kojih bi se s dovoljnom točnošću opisala zakonitost izlaznog signala. U okviru ove disertacije, a imajući u vidu trenutno stanje i trendove u području održavanja u industriji, istražena je mogućnost računalnog generiranja značajki ulaznih vibracijskih signala u cilju rane procjene nastanka i klasifikacije kvarova rotacijske opreme uslijed različitih tipova vibracijskih opterećenja. Pripremljen je i opisan eksperimentalni postav te kreiran i proveden eksperiment na temelju plana. Simulirano je osam različitih stanja opreme te su prikupljeni podaci za dvije različite brzine vrtnje. Modeliranje je provedeno primjenom algoritama dubokog strojnog učenja s ciljem točne klasifikacije tipa kvara, a na temelju automatski generiranih značajki vibracijskog signala. Tijekom procesa modeliranja, optimizirani su hiperparametri modela primjenom Bayesove optimizacije. Konačni rezultati dokazuju mogućnost primjene dubokih struktura neuronskih mreža za automatsko generiranje značajki ulaznog signala i klasifikaciju stanja opreme.

Ključne riječi: održavanje po stanju, automatizirani nadzor, vibracijska analiza, rotacijska oprema, duboko strojno učenje, konvolucijske neuronske mreže, klasifikacija stanja.

Summary

Recent trends focusing on Industry 4.0 concept arise a data-driven fault diagnosis as key topic in condition-based maintenance.

Traditional data-driven techniques of fault diagnosis require signal processing for feature extraction, as they are unable to work with raw signal data, consequently leading to need for expert knowledge and human work. In the context of this dissertation, considering the current state and trends in the field of maintenance in the industry, the possibility of computer-generated vibration signal input features was investigated in order to assess early on the occurrence and classification of rotational equipment failures due to different types of vibration loads. An experimental set-up was prepared and described, and a plan-based experiment was created and conducted. Eight different equipment conditions were simulated and data were collected for two different rotational speeds. The modeling was performed using deep machine learning algorithms with the aim of accurately classifying the fault type based on the automatically generated vibration signal features. During the modeling process, the hyperparameters of the model were optimized using Bayesian optimization. The final results prove the possibility of applying deep neural network structures to automatically generate input signal features and classify equipment states.

Keywords: condition based maintenance, monitoring, vibration analysis, rotary equipment, deep machine learning, convolution neural networks, states classification.

Podaci o mentoru

Dragutin Lisjak rođen je 28. studenog 1962. u Čakovcu gdje je završio osnovnu školu. Srednju školu završio je u Zagrebu. Diplomirao je 1989. godine na *Fakultetu strojarstva i brodogradnje Sveučilišta u Zagrebu*, usmjerenje Materijali. Do 1991. godine radio je u tvornici *Končar - Alati i strojevi* na poslovima tehnologije i konstrukcije izrade alata za tlačni lijev aluminija. 1991. godine zaposlio se kao znanstveni novak u *Zavodu za materijale FSB-a*. Magistarski rad pod naslovom *Razvoj ekspertnog sustava za rješavanje problema abrazijskog trošenja* obranio je na *Fakultetu strojarstva i brodogradnje Sveučilišta u Zagrebu*, 1998. godine, a disertaciju pod naslovom *Primjena metoda umjetne inteligencije pri izboru materijala* 2004. godine.

U znanstveno-nastavno zvanje docent izabran je 18. studenog 2008. na *Zavodu za industrijsko inženjerstvo, Fakulteta strojarstva i brodogradnje, Sveučilišta u Zagrebu*, gdje dalje nastavlja sa znanstveno- nastavnom aktivnošću.

Aktivno je sudjelovao u radu ili bio voditelj na slijedećim znanstvenim, tehnologijskim, informatičkim i projektima kratkoročne potpore istraživanju: *Računalne simulacije i razvoj materijala* (1997. - 2002.), *Razvoj materijala i procesa računalnim modeliranjem* (2000. - 2006.), *Difuzijsko modificiranje metala površinskim karbidnim slojevima* (2002. – 2006.), *Modeliranje svojstva materijala i parametara procesa* (2007. - 2011.-), *MAS-PLM TEMPUS JP: Master studies and continuing education network in product lifecycle management and sustainable production* (Project N° 144959-TEMPUS-2008-IT-JPCR) (2010. - 2012.), te *SUMAT - Razvoj i primjena suvremenih materijala* (2002. - 2004. godine). Bio je voditelj dvaju informatičkih projekata MZOŠ-a: *Identifikacija čelika prema kemijskom sastavu i standardnim oznakama* (2003.) i *Identifikacija aluminija i aluminijskih legura prema kemijskom sastavu i standardnim oznakama* (2004.) te 1 projekta PoC6 (HAMAG BICRO) *Inteligentni računalni sustav rane dijagnostike kvarova rotacijskih elemenata strojeva* (2016.). Trenutno sudjeluje na međunarodnom hrvatsko-kanadskom znanstvenom projektu financiranom od znanstvenog komiteta NATO STO AVT-SP- 001 *Development and evaluation of an advanced PACVD TiN/TiCN coating for military vehicle sliding bearing application* (2018. godina). Od projekata kratkoročne potpore istraživanju, voditelje je projekta *Tehnologija bežičnih senzorskih mreža u održavanju dislocirane strojarske opreme* (2015.), a sudjeluje i u radu

projekata *Napredni sustavi za kompenzaciju jalove snage mikromreže bušaćeg postrojenja za istraživanje ugljikovodika* (2016. godine) i *Razvoj i implementacija naprednih regulacijskih sustava za procese naftnog bušenja* (2018. godine). Također je bio voditelj internog fakultetskog projekta *Uvođenje informacijskog sustava održavanja InforEAM FSB* (2015. - 2018.).

Tijekom rada na *Zavodu za materijale* kontinuirano je sudjelovao u izvođenju nastave na dodiplomskom studiju i to iz sljedećih kolegija: vježbe iz kolegija *Materijali I, Materijali II, Materijali III, Materijali u brodogradnji, Materijali u zrakoplovstvu i Izbor materijala*.

Na *Zavodu za industrijsko inženjerstvo* na dodiplomskom studiju nositelj je kolegija: *Informacijski sustavi, Informatički menadžment, Održavanje, Održavanje u zrakoplovstvu, Dijagnostika u održavanju, Ekologija i održavanje, Pouzdanost tehničkih sustava te Višekriterijalno optimiranje i odlučivanje*. Na poslijediplomskom specijalističkom studiju nositelj je kolegija *Informacijski sustavi, Dijagnostika u održavanju, Menadžment održavanja i Ekologija i održavanje*. Na poslijediplomskom doktorskom studiju nositelj je kolegija *Inteligentni informacijski sustavi i Menadžment održavanja*.

Trenutno obnaša funkciju predstojnika *Zavoda za industrijsko inženjerstvo*.

Član je *Hrvatskog društva održavatelja, Hrvatskog društva za materijale i tribologiju, Američkog društva za materijale (ASM), te Američkog udruženja za umjetnu inteligenciju (AAAI)*.

Aktivno se služi engleskim, a pasivno njemačkim jezikom, oženjen je i otac dvojice sinova.

Sadržaj

1. UVOD.....	1
1.1 Motivacija.....	1
1.2 Cilj i hipoteza rada.....	4
1.3 Metodologija i plan istraživanja.....	4
1.3.1 Razvoj eksperimentalnog postava.....	4
1.3.2 Plan pokusa.....	5
1.3.3 Razvoj modela.....	5
1.4 Struktura rada.....	6
2. PREGLED POSTOJEĆEG STANJA.....	8
2.1 Ciljevi održavanja.....	8
2.2 Strategije održavanja.....	9
2.2.1 Korektivno održavanje.....	10
2.2.2 Plansko preventivno održavanje.....	12
2.2.3 Održavanje po stanju.....	13
2.3 Tehnike prediktivnog održavanja po stanju.....	17
2.3.1 Vibrodijagnostika.....	18
2.3.2 Dijagnostika akustičnom emisijom.....	20
2.4 Inteligentno prediktivno održavanje po stanju.....	21
2.4.1 Prikupljanje, obrada signala i određivanje značajki.....	23
2.4.2 Inteligentna dijagnostika stanja.....	25
2.5 Umjetne neuronske mreže.....	26
2.6 Duboke neuronske mreže.....	28
2.7 Primjena dubokog učenja u prediktivnom održavanju po stanju.....	30
2.8 Identifikacija i klasifikacija kvarova u ranoj fazi temeljena na dubokom učenju.....	32
2.9 Sažetak poglavlja.....	34
3. KONVOLUCIJSKE NEURONSKE MREŽE.....	36
3.1 Konvolucijski sloj i aktivacijska funkcija.....	36
3.2 Sloj sažimanja.....	42
3.3 Izlazni sloj.....	42
3.4 Učenje konvolucijske neuronske mreže.....	43

3.4.1	<i>Gradijentni spust</i>	44
3.4.2	<i>Algoritam povratnog prostiranja pogreške u konvolucijskim slojevima</i>	45
3.4.3	<i>Algoritam povratnog prostiranja pogreške u slojevima sažimanja</i>	49
3.5	Normalizacija grupe	49
3.6	Regularizacija	51
3.7	Određivanje hiperparametara modela	51
3.8	Računalne tehnike optimizacije hiperparametara modela	54
3.8.1	<i>Mrežno pretraživanje</i>	55
3.8.2	<i>Nasumično pretraživanje</i>	55
3.8.3	<i>Genetski algoritam</i>	55
3.8.4	<i>Bayesova optimizacija</i>	56
3.9	Sažetak poglavlja	61
4.	EKSPERIMENTALNI POSTAV I MJERNA OPREMA	63
4.1	Simulator kvarova rotacijske opreme.....	63
4.2	Mjerni sustav eksperimentalnog postava	65
4.2.1	<i>Beskontaktni optički senzor učestalosti vrtnje</i>	65
4.2.2	<i>Senzor akceleracije</i>	66
4.2.3	<i>Sustav za prikupljanje i obradu podataka</i>	67
4.2.4	<i>Korisničko sučelje mjernog računala</i>	68
5.	PLANIRANJE POKUSA I MJERENJE	71
5.1	Odabir modula simuliranih tipova uzroka kvarova	71
5.1.1	<i>Rotori za simuliranje neravnoteže</i>	72
5.1.2	<i>Ekscentričan rotor</i>	73
5.1.3	<i>Nagnuti rotor</i>	73
5.1.4	<i>Oštećeni kotrljajući ležajevi</i>	74
5.1.5	<i>Ostale korištene nadogradnje</i>	76
5.2	Izbor parametara eksperimenta.....	77
5.3	Snimanje i filtriranje signala	78
5.3.1	<i>Snimanje ispitnih uzoraka signala</i>	79
5.3.2	<i>Filtriranje signala niskopropusnim filterom</i>	79
6.	MODEL RANE PROCJENE UZROKA KVAROVA ROTACIJSKE OPREME ... 81	
6.1	Priprema podataka za učenje i testiranje modela	86

6.2	Struktura modela temeljenog na konvolucijskoj neuronskoj mreži.....	89
6.2.1	<i>Ulazni sloj mreže.....</i>	92
6.2.2	<i>Konvolucijski blokovi mreže.....</i>	93
6.2.3	<i>Slojevi sažimanja.....</i>	94
6.2.4	<i>Potpuno povezani sloj i SoftMax.....</i>	94
6.3	Učenje i optimizacija modela	95
6.3.1	<i>Stjecajna funkcija Bayesovog optimizatora.....</i>	97
6.3.2	<i>Odabir hiperparametara za optimizaciju.....</i>	99
6.4	Metrika za evaluaciju modela.....	103
6.5	Rezultati i evaluacija modela.....	105
6.5.1	<i>Model procjene kvarova na temelju signala vremenske domene.....</i>	105
6.5.2	<i>Model procjene kvarova na temelju signala frekvencijske domene.....</i>	122
6.5.3	<i>Dodatno testiranje razvijenih modela.....</i>	127
6.5.4	<i>Vizualizacija razvijenih modela.....</i>	128
6.6	Sažetak poglavlja	136
7.	ZAKLJUČAK.....	140
7.1	Osvrt na znanstvene doprinose i hipotezu istraživanja	140
7.2	Ograničenja provedenog istraživanja	141
7.3	Smjernice za daljnja istraživanja	142
8.	LITERATURA.....	144
	ŽIVOTOPIS	152
	BIOGRAPHY	153

Popis slika

Slika 2.1 Strategije održavanja prema [8].....	10
Slika 2.2 Princip održavanja po stanju [18]	14
Slika 2.3 Moderna podjela strategija održavanja [19].....	15
Slika 2.4 Pristupi prediktivnom održavanju po stanju [20]	16
Slika 2.5 Proces prediktivnog održavanja po stanju.....	18
Slika 2.6 Proces inteligentnog prediktivnog održavanja po stanju	22
Slika 2.7 Tehnologije inteligentne dijagnostike kvarova.....	25
Slika 3.1 Opća struktura konvolucijskih neuronskih mreža	37
Slika 3.2 Stvaranje mape značajki konvolucijom	38
Slika 3.3 Sigmoidalna aktivacijska funkcija	40
Slika 3.4 ReLU aktivacijska funkcija	41
Slika 3.5 Softmax funkcija.....	43
Slika 4.1 Eksperimentalni postav	63
Slika 4.2 Simulator kvarova.....	64
Slika 4.3 Ekran i optički senzor.....	66
Slika 4.4 Senzor akceleracije PCB 356B21	67
Slika 4.5 Razvijeno grafičko sučelje aplikacije za nadzor tijekom eksperimenta.....	69
Slika 4.6 Primjer strukture podataka u vremenskoj domeni	70
Slika 4.7 Primjer podataka u frekvencijskoj domeni.....	70
Slika 5.1 Rotor za simulaciju debalansa	72
Slika 5.2 Modul ekscentričnog rotora.....	73
Slika 5.3 Modul nagnutog rotora.....	74
Slika 5.4 Glavni elementi kotrljajućeg ležaja	75
Slika 5.5 Ležajevi s primarnim oštećenjima	76
Slika 5.6 Inercijski disk za vratilo	76
Slika 5.7 Opća funkcija Butterworth filtera trećeg reda	80
Slika 6.1 Proces istraživanja	82
Slika 6.2 Blok shema procesa razvoja modela za ranu procjenu uzroka kvara rotacijske opreme	86
Slika 6.3 Podjela podataka unutar svake klase uzroka kvara.....	87
Slika 6.4 Sadržaj pojedine datoteke ispitnog mjerenja.....	88
Slika 6.5 Blok shema skupova podataka za učenje i testiranje	89

Slika 6.6 Blok shema inicijalnog modela temeljenog na konvolucijskoj neuronskoj mreži sa ulaznim signalom u vremenskoj domeni (MVD)	90
Slika 6.7 Blok shema inicijalnog modela temeljenog na konvolucijskoj neuronskoj mreži sa ulaznim signalom u frekvencijskoj domeni (MFD)	91
Slika 6.8 Prezentacija RGB ulaznog signala mreže	92
Slika 6.9 Blok shema procesa učenja modela rane procjene tipa uzroka kvarova primjenom Bayesovog optimizatora.....	96
Slika 6.10 Blok shema pripreme za učenje optimizacijom parametara.....	100
Slika 6.11 Primjer matrice zabune.....	104
Slika 6.12 Trend funkcije cilja u procesu optimizacije	109
Slika 6.13 Grafikon minimalne i očekivane funkcije cilja tijekom procesa optimizacije ..	110
Slika 6.14 Grafikon rezultata uspješnosti klasifikacije UK, [%]	110
Slika 6.15 Grafikon grupiranih rezultata uspješnosti klasifikacije UK, [%]	111
Slika 6.16 Matrica zabune evaluacije najboljeg modela.....	112
Slika 6.17 Trend funkcije cilja tijekom dodatne optimizacije	116
Slika 6.18 Grafikon promjene vrijednosti minimalne i očekivane funkcije cilja tijekom dodatne optimizacije.....	117
Slika 6.19 Grafikon rezultata UK, [%] - dodatna optimizacija.....	118
Slika 6.20 Grafikon grupiranih rezultata UK[%] – dodatna optimizacija	118
Slika 6.21 Grafikon usporedbe rezultata provedenih optimizacija	119
Slika 6.22 Matrica zabune najuspješnijeg modela vremenske domene.....	120
Slika 6.23 Grafički prikaz karakteristične FFT vibracijske slike za klasu 1000NS.....	124
Slika 6.24 Grafički prikaz karakteristične FFT vibracijske slike za klasu 1000IMRF	124
Slika 6.25 Proces testiranja naučenog MVD/MFD modela	127

Popis tablica

Tablica 2.1 Primjena obrade signala i izdvajanja značajki u održavanju	24
Tablica 2.2 Literatura – neuronske mreže u prediktivnom održavanju po stanju	27
Tablica 3.1 Standardni pseudokod Bayesopt algoritma	58
Tablica 4.1 Osnovne tehničke karakteristike simulatora kvarova.....	65
Tablica 4.2 Osnovne tehničke karakteristike senzora akceleracije	67
Tablica 4.3 Tehničke karakteristike NI 9234 modula	68
Tablica 5.1 Parametri stanja eksperimenta	78
Tablica 6.1 Korišteni nadogradni moduli za simulaciju stanja opreme.....	86
Tablica 6.2 Dimenzije ulaznog signala MVD modela.....	93
Tablica 6.3 Dimenzije ulaznog signala MFD modela	93
Tablica 6.4 Rasponi optimizacijskih varijabli Bayesovog optimizatora	101
Tablica 6.5 Izračun hiperparametara preko optimizacijskih varijabli	102
Tablica 6.6 Hiperparametri Bayesovog optimizatora.....	102
Tablica 6.7 Iteracije Bayesove optimizacije MVD	106
Tablica 6.8 Najboljih 6 % kombinacija vrijednosti varijabli Bayesove optimizacije	111
Tablica 6.9 Prošireni rasponi optimizacijskih varijabli Bayesovog optimizatora.....	113
Tablica 6.10 Iteracije dodatne Bayesove optimizacije MVD	114
Tablica 6.12 Struktura optimalnog modela vremenske domene temeljenog na dubokoj konvolucijskoj neuronskoj mreži	121
Tablica 6.13 Rezultati optimizacije hiperparametara MFD.....	125
Tablica 6.14 Uspješnost klasifikacije MFD na evaluacijskom skupu podataka	126
Tablica 6.15 Rezultati testiranja MVD i MFD modela	128
Tablica 6.16 Broj aktivacija pojedinih slojeva optimalnog MVD	129
Tablica 6.17 Vizualizacija značajki MVD primjenom t-SNE algoritma.....	130
Tablica 6.18 Vizualizacija značajki MFD primjenom t-SNE algoritma	134
Tablica 6.19 Optimizacijske varijable najboljeg MVD prve Bayesove optimizacije	137
Tablica 6.20 Optimizacijske varijable najboljeg MVD	138

Popis oznaka

Svi korišteni simboli i oznake objašnjene su unutar teksta na mjestu pojavljivanja.

Popis kratica

Kratika	Opis
ADAM	Adaptivna procjena momentuma (engl. <i>adaptive moment estimation</i>)
AE	Akustična emisija (engl. <i>acoustic emission</i>)
AI	Umjetna inteligencija (engl. <i>artificial intelligence</i>)
ARD	Funkcija automatskog određivanja relevantosti (engl. <i>automatic relevance determination</i>)
BBF	Oštećenje kotrljajućeg elementa kotrljajućeg ležaja
BP	Algoritam povratnog prostiranja pogreške (engl. <i>back-propagation algorithm</i>)
BPNN	Unaprijedne mreže sa povratnim prostiranjem pogreške (engl. <i>back-propagation neural networks</i>)
CBF	Kombinirano oštećenje kotrljajućeg ležaja
CBM	Održavanje po stanju (engl. <i>condition based maintenance</i>)
CC	Računarstvo u oblacima (engl. <i>cloud computing</i>)
CM	Korektivno održavanje (engl. <i>corrective maintenance</i>),
CNN	Konvolucijske neuronske mreže (engl. <i>convolutional neural networks</i>)
CPS	Kibernetско-fizički sustav (engl. <i>cyber-physical system</i>)
CRF	Nagnutnost rotora kao uzrok kvara
DBM	Duboki Boltzmannovi strojevi (engl. <i>deep Boltzmann machine</i>)
DBN	Bayesove mreže (engl. <i>deep belief Bayes networks</i>)
EEMD	Empirijsko razlaganje skupa signala (engl. <i>ensemble empirical mode decompositon</i>)
EI	Očekivano poboljšanje (engl. <i>expected improvement</i>)
ERF	Ekscentričnost rotora kao uzrok kvara
FFT	Brza Fourierova pretvorba (engl. <i>fast Fourier transform</i>)

GCL	Geometrijska os rotacije (engl. <i>geometric centerline</i>)
GP	Gaussov slučajni proces (engl. <i>Gaussian process</i>)
GPU	Grafička procesorska jedinica (engl. <i>graphical processing unit</i>)
ICT	Informacijske i komunikacijske tehnologije (engl. <i>information and communications technology</i>)
IEPE	Senzor sa integriranim elektroničkim komponentama (engl. <i>integrated electronic piezoelectric</i>)
IIR	Beskonačni impulsni odziv (engl. <i>infinite impulse response</i>)
IMRF	Neravnoteža rotora kao uzrok kvara
IoT	Internet stvari (engl. <i>internet of things</i>)
IRBF	Oštećenje unutarnje staze kotrljanja kotrljajućeg ležaja
KB1	Konvolucijski blok tipa 1
KB2	Konvolucijski blok tipa 2
KB3	Konvolucijski blok tipa 3
LSTM	Mreža s dugom kratkoročnom memorijom (engl. <i>long short-term memory</i>)
M2M	Komunikacija između strojeva (engl. <i>machine-to-machine</i>)
MFD	Model frekvencijske domene
MVD	Model vremenske domene
NFN	Neuro-fuzzy mreže (engl. <i>neuro-fuzzy networks</i>)
OEE	Ukupna učinkovitost opreme (engl. <i>overall equipment effectiveness</i>)
ORBF	Oštećenje vanjske staze kotrljanja kotrljajućeg ležaja
PdM	Prediktivno održavanje (engl. <i>predictive maintenance</i>)
PIA	Glavna os inercije (engl. <i>principle inertia axis</i>)
PM	Plansko preventivno održavanje (engl. <i>preventive maintenance</i>)

RBFNN	Neuronske mreže sa radijalnim baznim funkcijama (engl. <i>radial basis function neural networks</i>)
RBM	Ograničeni Boltzmanovi strojevi (engl. <i>restricted Boltzmann machines</i>)
ReLU	Zglobna aktivacijska funkcija (engl. <i>rectifier linear unit</i>)
RFID	Radiofrekvencijska identifikacija (engl. <i>radio-frequency identification</i>)
RMS	Korijen srednje vrijednosti kvadrata (engl. <i>root mean square</i>)
RMSProp	Algoritam povratnog prostiranja korijena srednje vrijednosti kvadrata pogreške (engl. <i>root mean square propagation</i>)
RNN	Povratne neuronske mreže (engl. <i>recurrent neural networks</i>)
SAE	Naslagani autoenkoderi (engl. <i>stacked autoencoders</i>)
SOM	Samoorganizirajuće mape (engl. <i>self-organizing maps</i>)
SPM	Metoda udarnih impulsa (engl. <i>shock pulse method</i>)
SvMR	Regresija realizirana pomoću metode potpornih vektora (engl. <i>support vector machines regression</i>)
TPM	Cjelovito produktivno održavanje (engl. <i>total productive maintenance</i>)
t-SNE	Algoritam za smanjenje broja dimenzija t-SNE (engl. <i>t-distributed stochastic neighbor embedding</i>)

1. UVOD

Razvojem informacijske i komunikacijske tehnologije (engl. *Information and communications technology*, ICT) i integracijom analitike u računalom podržane procese, industrija se suočava sa novim prilikama, ali i izazovima u potrebi da uravnoteži vlastite kompetencije i potrebe tržišta.

Porast svijesti o mogućnostima primjene interneta stvari (engl. *internet of things*, IOT) i koncepta Industrije 4.0, kako na globalnoj, tako i na lokalnoj razini, omogućuje međusobno povezivanje sve većeg broja objekata te pristup podacima koji su povezani s tim objektima, npr. podacima sa senzora ili izvještajima. Sam pristup velikoj količini podataka ne znači nužno i ostvarivanje koristi od samih podataka, ali otvara niz mogućnosti na koje se ti podaci mogu iskoristiti. Uz pretpostavku o potrebi održavanja spomenutih objekata, moguća je primjena strategije prediktivnog održavanja. Prediktivno održavanje pripada disciplini održavanja po stanju (engl. *condition based maintenance*, CBM) uz pomoć koje je moguće, uz poznavanje trenutnih značajki realnog sustava odnosno stanja opreme, procijeniti njezino trenutno stanje, a s ciljem prevencije fatalnih kvarova i utjecaja na njenu pouzdanost.

Uvodni dio rada objašnjava izazove područja održavanja rotacijske opreme po stanju. Dan je pregled prošlih istraživanja kako bi se objasnila motivacija za istraživanje u ovoj disertaciji. Postavljen je cilj rada i hipoteza rada s očekivanim znanstvenim doprinosima. Objasnjena je korištena terminologija i prikazana je struktura disertacije.

1.1 Motivacija

Prema [1], održavanje opreme može sačinjavati 10 – 40 % udjela proizvodnih troškova, autori u [2] definiraju udio održavanja u rasponu 15 – 70 %, dok prema [3] troškovi održavanja mogu činiti 15 – 60 % ukupne cijene proizvodnje pojedinog proizvoda. Nadalje, smatra se [4] kako je 1/3 troškova održavanja posljedica lošeg planiranja zahvata, loše primjene preventivnog održavanja i sl., a istraživanje provedeno u [5] navodi kako čak 70 % ispitanih smatra funkciju održavanja troškovnim centrom unutar proizvodnog poduzeća. Iz navedenog, može se zaključiti kako troškovi održavanja mogu imati značajan udio u troškovima proizvodnje. Dodatno, može se pretpostaviti kako ulaganje u razvoj funkcije održavanja i implementacija novih tehnologija održavanja može imati utjecaja na smanjenje udjela troškova održavanja u ukupnim troškovima proizvodnje.

Nakon pojave koncepta cjelovitog produktivnog održavanja (engl. *total productive maintenance*, TPM) početkom 1950.-tih [6], održavanje kao područje počinje poprimati, u odnosu na dotadašnju usmjerenost na organizaciju i kvalitetu, značajke područja usmjerenog ka korisniku odnosno stvaranju vrijednosti. 1970.-tih godina se, u specifičnim industrijama (zrakoplovstvo), počinje primjenjivati strategija održavanja opreme temeljena na trenutnim mjerenjima i CBM-u, koje se otada dalje razvija i zahvaća sve veći broj ostalih industrija. U posljednjih 30-tak godina, provedena su brojna istraživanja u području primjene CBM-a u raznim industrijama, pa je tako danas moguće tvrditi kako strategija održavanja po stanju s ekonomske strane pozitivno utječe na funkciju održavanja.

Povijesno gledano, industrijska proizvodnja se u ljudskoj civilizaciji razvija kroz zadnja dva stoljeća, a najveći napredak dogodio se nakon otkrića i inovacija koje su u potpunosti mijenjali smjer razmišljanja i funkcioniranja tadašnjih industrijskih postrojenja. Takvi zaokreti nazivaju se industrijske revolucije, a dosada su zabilježena tri takva događaja u povijesti industrijskog razvoja. Razvoj industrijske proizvodnje može se povezati sa razvojem strategija održavanja, koje su nastale kao posljedica potreba za efikasnijom upotrebom jednog od najvažnijih proizvodnih resursa – proizvodne opreme. Tako se održavanje razvijalo uz industrijske revolucije:

- Industrija 1.0: Reaktivno održavanje. Na održavanje se nisu trošili dostupni resursi, te se prakticirala strategija rada stroja do kvara. Bez ikakvih troškova održavanja i radne snage koja se time bavila, vrijeme nedostupnosti strojeva je bilo veliko.
- Industrija 2.0: Preventivno održavanje. Periodično održavanje je uključivalo mijenjanje komponenti prije pojave kvara, a intervali održavanja su određivani iskustveno ili na preporuku proizvođača opreme.
- Industrija 3.0: Prediktivno održavanje. Počinje implementacija održavanja po stanju u realnom vremenu, praćenjem stanja opreme u operaciji. Granice poslije kojih se proglašava vrijeme za održavanje su određene statističkim i iskustvenim mjerilima.

Relevantni industrijski stručnjaci i znanstvenici slažu se u međusobnim istraživanjima da je u tijeku začetak četvrte industrijske revolucije.

Ubrzanim napretkom ICT-a u zadnjih nekoliko godina, olakšana je implementacija senzora, opreme za prikupljanje podataka, bežično upravljanih uređaja i računalnih rješenja za udaljeni nadzor industrijskih okruženja. Integracija nabrojanih tehnologija sa opremom instaliranom u industriji kombinirana sa rješenjima za obradu i analizu podataka može se nazvati kibernetičko-

fizički sustav (engl. *cyber-physical system*, CPS) [7], čime se stvara fizički sustav čije su operacije nadzirane, koordinirane, upravljane i integrirane računalnom i komunikacijskom podrškom. CPS je predstavljen kao ključna tehnologija za primjenu koncepta Industrije 4.0, koji će prema nekim istraživanjima [8,9] imati značajan utjecaj na industriju i gospodarstvo (u Njemačkoj se primjenom Industrije 4.0 očekuje rast produktivnosti od 5 – 8 % uz uključene troškove konverzije). Na sajmu u Hannoveru (*Hannover Messe*) održanom 2016. godine, *Prediktivno održavanje 4.0* (engl. *predictive maintenance 4.0*) predstavljeno je kao ključna tema koncepta Industrije 4.0. Rastući broj primjena koncepta industrije 4.0 u vidu istraživačkih i pilot projekata generira prostor za primjenu prediktivnog održavanja po stanju integriranog u 4.0 koncept.

Industrija 4.0 usko je povezana i s nekim drugim konceptima, kao što su komunikacija između strojeva (engl. *machine to machine*, M2M) [10], tehnologija radiofrekvencijske identifikacije (engl. *radio-frequency identification*, RFID), IOT, računarstvo u oblacima (engl. *cloud computing*, CC), kao i sustavima za podršku donošenju odluka temeljenim na dubinskoj analizi podataka i umjetnoj inteligenciji (engl. *artificial intelligence*, AI).

U tvornici vođenoj prema konceptu Industrije 4.0, strojevi su međusobno povezani na način koji omogućuje sustavno prikupljanje, razmjenu i analizu podataka. Takav pristup, u konačnici, kombinira prednosti industrijske proizvodnje s internetskim tehnologijama i značajno mijenja proizvodni proces, strategije održavanja i upravljanje održavanjem. Shvativši potencijal, različita poduzeća razvila su personalizirane pristupe i vlastite strategije, dok druga još uvijek nisu sigurna trebaju li i u kojem trenutku i obimu krenuti s implementacijom koncepta.

Promatrano iz aspekta održavanja, može se reći kako se sustav koji može samostalno procijeniti vlastito stanje i kvar u ranoj fazi i pritom pri procjeni koristiti povijesne podatke vlastitog sustava za prikupljanje podataka i podatke sličnih sustava te na temelju procjene donositi odluke o aktivnostima održavanja može smatrati *samosvjesnim*.

Takva procjena može biti donesena na temelju dubinske analize podataka prikupljenih na odgovarajućoj opremi i njenom okruženju. Iako se današnje analitičke mogućnosti razvijaju, i dalje je postojeća praznina između ostvarenja vizija koncepta Industrije 4.0 i sustavnih istraživanja, koja bi kao rezultat pružila novo iskustvo za implementaciju računalne inteligencije u održavanju.

1.2 Cilj i hipoteza rada

Uzevši u obzir postojeće trendove razvoja u području prediktivnog održavanja i u prethodnom odjeljku definiranu istraživačku prazninu, ciljevi ovog istraživanja predstavljaju mogućnosti primjene algoritama dubokog strojnog učenja za procjenu uzroka kvarova rotacijske opreme te razvoj modela za računalno generiranje utjecajnih značajki ulaznih vibracijskih signala za predefinirane tipove vibracijskih opterećenja. Konačno, na temelju računalno generiranih utjecajnih značajki razvit će se model za ranu procjenu kvarova rotacijske opreme.

U tom kontekstu moguće je postaviti sljedeću hipotezu istraživanja:

Moguće je na temelju ulaznih vibracijskih signala te predefiniranih vibracijskih opterećenja, računalno generirati utjecajne značajke te njihovom implementacijom u novorazvijeni računalni model dubokog strojnog učenja, procijeniti potencijalne kvarove rotacijske opreme u ranoj fazi njihova nastanka.

Definirani znanstveni doprinosi ovog istraživanja su:

1. razvijen model računalnog generiranja utjecajnih značajki vibracijskog signala temeljen na analizi velike količine ulaznih vibracijskih signala, te
2. razvijen računalni model dubokog strojnog učenja rane procjene kvarova rotacijske opreme temeljen na računalno generiranim utjecajnim značajkama vibracijskih signala.

1.3 Metodologija i plan istraživanja

Planirano eksperimentalno istraživanje provedeno je u nekoliko faza opisanih u nastavku.

1.3.1 Razvoj eksperimentalnog postava

U cilju strukturiranja i verifikacije modela za procjenu kvarova rotacijske opreme, u prvoj fazi istraživanja proveden je razvoj i izrada eksperimentalnog ispitnog postava. On je temeljen na postojećoj konstrukciji simulatora kvarova rotacijske opreme smještenog unutar *Laboratorija za održavanje FSB-a*. Simulator kvarova rotacijske opreme jest modularni uređaj sastavljen u početnoj konfiguraciji od vratila oslonjenog na dva kuglična ležaja te preko spojke pogonjenog trofaznim kaveznim asinkronim motorom upravljanim frekvencijskim pretvornikom opremljenim regulatorom brzine vrtnje. Na predmetnom ispitnom postavu su za određeno predefinirano vibracijsko opterećenje potom ugrađivani moduli za simulaciju pojedinih tipova kvarova (debalans, nagnuti rotor, ekscentrični rotor, kvar ležaja). Mjerni sustav ispitnog postava sastoji se od senzora akceleracije s mjerenjem u tri osi smještenih na položajima glavnih osi,

senzora brzine vrtnje rotora te sustava za prikupljanje podataka spojenog na upravljačko računalo. Signali sa senzora prikupljeni su akvizicijskim sustavom *National Instruments* i pohranjivani na mjesto za pohranu u prikladnom formatu za daljnju obradu.

Prikupljeni podaci obrađivani su računalom konfiguriranim za obradu podataka i modeliranje na grafičkoj kartici sa grafičkim procesorima serije *Turing (GeForce RTX 2070)*.

1.3.2 Plan pokusa

Pri planiranju pokusa, određena su mjesta postavljanja senzora akceleracije i parametri rada simulatora tokom mjerenja, načini simulacije pojedinih tipova kvarova te parametri sustava za prikupljanje i pohranu podataka. Parametri rada simulatora odabrani su s obzirom na preporuke proizvođača, značajke simuliranog kvara i na temelju iskustva, a različite kombinacije parametara određene su eksperimentalno. Tijekom eksperimenta, korišteni su unaprijed pripremljeni nadogradni moduli simulatora kvarova, pomoću kojih su simulirani različiti karakteristični kvarovi rotacijske opreme.

Snimanje vibracijskog stanja provedeno je pomoću senzora akceleracije mjerenjem u tri osi, a u svim eksperimentima korištena je *National Instruments* oprema za prikupljanje podataka. Sam postupak mjerenja slijedi smjernice dane normom *ISO 20816-1:2016* [11], a rezultati mjerenja amplitude akceleracije pohranjivani su u vremenskoj i frekvencijskoj domeni. Za svaki simulirani tip kvara provodi se 1 500 mjerenja pri brzini vrtnje motora od 1 000 o/min i 1 500 o/min, a frekvencija prikupljanja vibracijskog signala iznosi 51,2 kHz te je u svakom mjerenju prikupljeno 25 600 zapisa o amplitudama akceleracija senzora u tri smjera (X, Yi Z), a koji su dodatno obilježeni trenutnim podacima o parametrima rada simulatora, vremenskim identifikatorom te podatkom o trenutno simuliranom stanju.

Prikupljeni i pohranjeni zapisi iskorišteni su za razvoj modela za računalno generiranje utjecajnih značajnih ulaznih vibracijskih signala za predefinirane tipove kvarova rotacijske opreme te razvoj računalnog modela dubokog strojnog učenja rane procjenu kvarova.

1.3.3 Razvoj modela

Promjena vibracijskog obrasca posljedica je promjene stanja tj. degradacije funkcionalnosti opreme. Pojava određenog tipa kvara već u ranoj fazi rezultira promjenom vibracijskog obrasca, pa ako se pretpostavi unikatni vibracijski obrazac za svaki tip kvara, pretpostavlja se da je moguće izraditi računalni model rane procjene kvarova rotacijske opreme temeljene na podacima vibracijskih obrazaca. Postojeći klasični modeli strojnog učenja za procjenu kvarova rotacijske opreme za operacije pretprocesiranja te izdvajanja relevantnih značajki ulaznih

vibracijskih signala iziskuju značajni utrošak vremena te posjedovanje ekspertnog znanja. Kako bi se navedeno izbjeglo, razvijeni su algoritmi dubokog strojnog učenja temeljeni na konvolucijskim umjetnim neuronskim mrežama čiji se proces učenja temelji na računalno generiranim značajkama ulaznog vibracijskog signala u vremenskoj domeni. Razvojem algoritma smatra se postupak identifikacije i primjene optimalne arhitekture konvolucijske neuronske mreže- Za modeliranje je primijenjen programski paket *MATLAB*.

Prikupljeni podaci stohastički su podijeljeni na skupove za učenje, validaciju i testiranje modela, a početni hiperparametri učenja odabrani na temelju preporuka iz literature. Treniranje modela dubokog strojnog učenja izvođeno je unaprijed s povratnim prostiranjem pogreške u više epoha (iteracija) učenja, uz primjenu gradijentnog postupka traženja minimuma kriterijske funkcije. Ocjena kvalitete dobivenog modela određivana je prema kriterijima uspješnosti klasifikacije tipa kvara na definiranim testnim skupovima. Optimizacija hiperparametara provedit će se Bayesovom optimizacijom.

1.4 Struktura rada

Ova disertacija podijeljena je u nekoliko povezanih cjelina. Uvodni dio rada donosi opće informacije o području održavanja opreme te opisuje ključne pojmove održavanja po stanju, što omogućuje shvaćanje šire slike i položaj područja u kojem se provodi istraživanje. Nadalje, uvodni dio daje pregled temeljnih korištenih tehnika za održavanje po stanju te naglašava važnost primjene vibrodijagnostike. U drugom dijelu, opisano je područje inteligentnog održavanja po stanju te je dan literaturni pregled trenutno aktualnih arhitektura i tehnika dubokog učenja u području inteligentne dijagnostike stanja i identifikacije uzroka kvarova, a za tehnike modeliranja opisani su korišteni modeli. Sljedeće poglavlje definira i detaljno opisuje konvolucijske neuronske mreže, njihovu arhitekturu i slojeve te u koracima objašnjava princip učenja istih. Nadalje, odjeljak 3.4 opisuje proces učenja konvolucijske neuronske mreže. Dodatno, na temelju izvedenih zaključaka o važnosti odabira ispravnih vrijednosti hiperparametara, navedene su i ukratko definirane osnovne tehnike optimizacije hiperparametara. Objašnjena je ispravnost korištenja Bayesove optimizacije za potrebe optimizacije hiperparametara konvolucijskih neuronskih mreža te je navedena tehnika i detaljno opisana.

Poglavlje 4 detaljno opisuje laboratorijsku i mjernu opremu korištenu pri provedbi eksperimentalnog dijela rada.

U petom poglavlju opisani su simulirani tipovi kvarova kao i korišteni nadogradni moduli potrebni za simulaciju pojedinih tipova uzroka kvarova. Opisani su korišteni parametri izvođenja glavnog eksperimenta te način provedbe mjerenja i filtriranja signala.

U šestom poglavlju opisan je razvoj računalnih modela za procjenu stanja rotacijske opreme te rezultati testiranja modela. U prvom dijelu, opisan je korišteni postupak za razvoj modela, definirani su koraci i korištene tehnike. Nadalje, izneseni su rezultati uspješnosti klasifikacije razvijenih modela na evaluacijskim i testnim skupovima.

Na temelju ostvarenih rezultata doneseni su zaključci rada.

2. PREGLED POSTOJEĆEG STANJA

Upravljanje funkcijom održavanja opreme unutar industrijskih postrojenja važan je aspekt u uspostavljanju ispravne, funkcionalne i profitabilne industrijske cjeline. Održavanje se može definirati kao skup svih tehničkih i odgovarajućih administrativnih aktivnosti predviđenih za očuvanje nekog sredstva rada tj. radnog sustava ili dovođenje istog u stanje u kojem on može obavljati predviđenu funkciju [12].

2.1 Ciljevi održavanja

Početak 20. stoljeća, održavanje se smatralo dodatnim troškom proizvodnog sustava, nije povećavalo vrijednost proizvoda te se svodilo na popravak ili zamjenu komponenti u slučaju kvara. Tijekom i nakon Drugog svjetskog rata, razvojem inženjerstva i znanstvenog pristupa tehnologiji, razvijeni su i drugi oblici održavanja, koji su se pokazali troškovno povoljnijima i adekvatnijim za primjenu u proizvodnoj okolini (npr. preventivno održavanje). U današnje vrijeme povećana svijest o pitanjima kao što su sigurnost, okolišni utjecaji, kvaliteta proizvoda i usluga čini održavanje jednom od najvažnijih funkcija koje doprinose uspjehu industrije i stvaranju konkurentske prednosti.

Ciljevi održavanja trebaju biti u skladu s proizvodnim ciljevima i njima podređeni. Odnos između ciljeva održavanja i proizvodnje reflektira se u kontinuiranoj proizvodnji i održavanju postrojenja u najboljem mogućem stanju. Tipično, ciljevi održavanja mogu se svrstati u tri skupine [4]:

- Tehnički ciljevi: operativni imperativi vezani za zadovoljavajuću razinu dostupnosti opreme i sigurnosti ljudi. Općenito prihvaćena metoda za mjerenje ispunjenja ciljeva je ukupna učinkovitost opreme (engl. *overall equipment effectiveness*, OEE).
- Pravni ciljevi / prisilni propis: potrebno je ispuniti sve važeće propise za električne uređaje, opremu pod tlakom, vozila, zaštitna sredstva, itd.
- Financijski ciljevi: uključuju termin troškova u održavanju, tj. potrebu da se funkcija održavanja obavlja sa minimalnim troškovima.

Općenito, ciljevi procesa održavanja se mogu po važnosti poredati na sljedeći način [4]:

- 1) maksimiziranje proizvodnje ili povećanje raspoloživosti sa minimalnom cijenom, a uz odgovarajuću kvalitetu i sigurnost,
- 2) smanjenje kvarova i neočekivanih ispada,
- 3) optimiranje korištenja resursa,

- 4) smanjenje vremena u kvaru,
- 5) poboljšanje kontrole stanja zaliha rezervnih dijelova,
- 6) povećanje učinkovitosti opreme i smanjenje stope škarta,
- 7) smanjenje potrošnje energije,
- 8) optimizacija vijeka trajanja opreme,
- 9) pouzdana kontrola troškova održavanja, te
- 10) prepoznavanje mogućnosti i provedba smanjenja troškova.

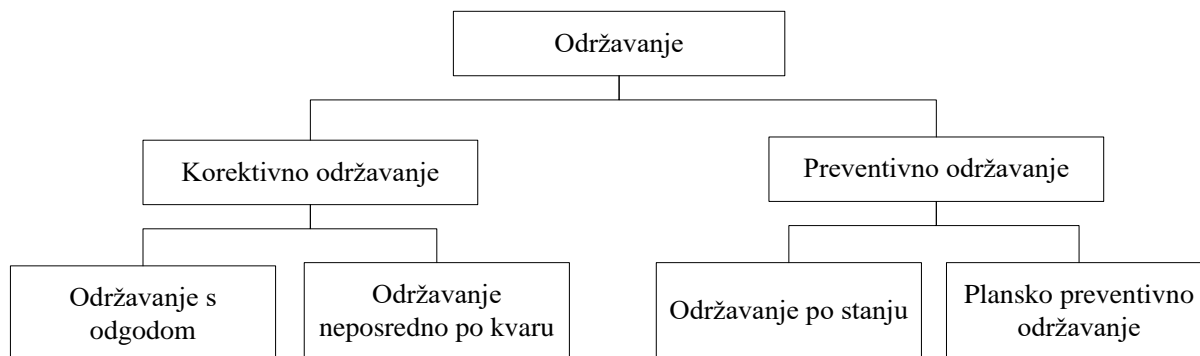
Zadaci održavanja mogu uključivati niz aktivnosti održavanja: pregled, nadzor, redovno održavanje, remont, obnovu, popravak i dr. Nadzor je vrsta aktivnosti koje se izvode ručno ili automatski, kontinuirano ili povremeno s namjerom dobivanja stvarnog stanja opreme, koja se može koristiti za procjenu promjene parametara.

Održavanje kroz kontinuiran nadzor stanja opreme pripada strategiji održavanja po stanju, prepoznate kao najučinkovitije strategije održavanja u širokom spektru različitih industrija. U početku, oprema se održavala po pojavi kvara, što je osiguravalo maksimalno vrijeme između zastoja, no često s katastrofalnim posljedicama po sigurnost, proizvodne gubitke i troškove održavanja. Rješenje navedenih problema ponudila je strategija preventivnog održavanja, prema kojoj se aktivnosti održavanja odvijaju u predefiniranim intervalima te se na taj način smanjuje vjerojatnost katastrofalnih posljedica kvarova. S druge strane, primjena takve strategije znači i povećani utrošak doknadnih dijelova, kao i veći intenzitet aktivnosti održavanja.

U narednim odjeljcima, bit će opisane strategije održavanja, a detaljnije će se objasniti strategija održavanja po stanju. Vibracijska analiza, kao temeljna i najvažnija tehnika održavanja po stanju rotacijske opreme opisana je u odjeljku 2.3.1. Nadalje, u pregled dosadašnjih spoznaja uključeno je područje dubokog strojnog učenja s naglaskom na identifikaciju kvarova i tipova kvarova u ranoj fazi.

2.2 Strategije održavanja

Autori različito dijele strategije održavanja, no u osnovi se one mogu podijeliti na reaktivne i preventivne strategije. Općenita taksonomija i podjela strategije održavanja prikazana je grafički (vidi sliku 2.1).



Slika 2.1 Strategije održavanja prema [8]

Pristup temeljen na obavljanju aktivnosti održavanja po pojavi ili pronalasku kvara definira se kao korektivno održavanje, dok se osnovna filozofija koja stoji u pozadini preventivnog održavanja svodi se na izvođenje zahvata prije same pojave kvara, a unutar koje se razlikuju 2 osnovna načina održavanja:

- a) pristup temeljen na određivanju konstantnog ciklusa nakon kojeg će biti provedene aktivnosti održavanja,
- b) održavanje temeljeno na procjeni stvarnog stanja opreme i potrebi za održavanjem.

2.2.1 Korektivno održavanje

Korektivno održavanje (engl. *corrective maintenance*, CM) jest tradicionalni pristup prema kojem se rad opreme nastavlja sve do pojave kvara, a aktivnosti održavanja provode tek po pojavi kvara, što se smatra prihvatljivim samo u slučajevima:

- kad su takvi kvarovi dijagnostički nezahtjevni i vrijeme popravka kratko,
- kad ne postoji zahtjev za visokom raspoloživošću opreme, pa vrijeme potrebno za identifikaciju kvara i popravak nema značajnu ulogu u proizvodnim troškovima, koji nastaju zbog zastoja stroja.

U nekim slučajevima kvar nije moguće predvidjeti pomoću instrumenata ili analize te je moguće kvar otkloniti samo nakon nastanka. U takvim slučajevima, nakon kvara, zadatak održavanja je da se stroj vrati u stanje u kojem može obavljati tražene funkcije što je brže moguće [13]. Dakle, CM je najbolje koristiti u ne-kritičnim područjima gdje su kapitalni troškovi mali, posljedice kvara blage, nema neposrednih sigurnosnih rizika, a moguća je brza identifikacija kvara i vraćanje u operativno stanje.

Niu i dr. [14] dodatno dijele strategiju CM-a prema trenutku odvijanja aktivnosti održavanja, pa tako razlikuju održavanje neposredno nakon pojave kvara te održavanje s odgođenim radovima. U načelu, takva strategija održavanja osigurava najduže vrijeme između aktivnosti

održavanja, no posljedice kvarova mogu biti katastrofalne i rezultirati ozbiljnom štetom ne samo na opremi, već u okruženju u kojem je oprema instalirana. Odjel održavanja kontinuirano djeluje na način da je spreman za neplanirane kvarove, osiguravajući u svakom trenutku potrebnu količinu doknadnih dijelova i ljudskih resursa za pravovremenu i brzu reakciju. No, priroda kvarova u slučaju korektivnog održavanja češće uvjetuje povećane aktivnosti održavanja, pa su potrebni i dodatni doknadni dijelovi i vrijeme. U takvim slučajevima, većinu troškova čini upravo trošak zastoja odnosno trošak vremena u kojem oprema ne radi, koji može i premašiti samu vrijednost opreme. Ostale troškove primjene takve strategije čine troškovi skladišta doknadnih dijelova, kao i troškovi prekovremenih sati održavatelja. Dodatno, potencijalno veći trošak u budućem razdoblju može se uzrokovati ugradnjom nekvalitetnih (troškovno povoljnijih) doknadnih dijelova. Korektivno održavanje najneučinkovitija je strategija održavanja opreme, no i kao takva može naći svoju primjenu u nekim industrijama. Izdvojiti se mogu industrije sa većim brojem manjih strojeva (npr. tekstilna industrija sa strojevima za šivanje), gdje gubitak jednog stroja nije kritičan za proizvodni proces, a kvarovi rijetko mogu imati ozbiljne posljedice te mogu biti jednostavno i brzo identificirani i uklonjeni. Korektivno održavanje ima svoje prednosti. Planiranje je vrlo jednostavno, jer je potrebno održavanje samo u slučaju kvara, pa se pri planiranju u obzir uzima samo vjerojatnost kvara. Zadaci održavanja nisu planirani sve dok ono nije stvarno potrebno.

Ukoliko je neprimjereno primijenjena kao strategija održavanja, nedostaci korektivnog održavanja su [13] :

- kvar će se dogoditi u nepredviđeno vrijeme, npr. kad je oprema u punom opterećenju ili pri pokretanju,
- kvar komponente može proći nezamijećen, a posljedično voditi skupljem oštećenju, npr. kvar ležaja može uzrokovati oštećenje vratila,
- mogu se očekivati opasne i skupe posljedice kvarova,
- nema dostupnih podataka o prošlom, sadašnjem ni budućem stanju opreme,
- najčešće je potreban veći tim održavatelja *na čekanju*,
- potrebno je osigurati zalihu rezervnih dijelova, da bi se popravak izvršio u što kraćem roku, te
- kvarovi koji nadilaze kapacitete tima održavatelja vode prema nepotpunim popravcima i *gašenju požara*, odnosno privremenom otklanjanju kvarova samo da bi oprema barem nakratko bila u operativnom stanju.

Unatoč tehnološkom napretku moderne ere, i dalje nije neobično pronaći i velika industrijska postrojenja i opremu, koja se održavaju po strategiji korektivnog održavanja

2.2.2 Plansko preventivno održavanje

Plansko preventivno održavanje (engl. *preventive maintenance*, PM) razvojno se nastavlja na strategiju korektivnog održavanja. Primjenom ovakve strategije, preventivne aktivnosti održavanja obavljaju se prema definiranim konstantnim ciklusima, uzimajući pritom u obzir očekivano i unaprijed definirano vrijeme između kvarova (engl. *time between failures*), tj. aktivnosti održavanja obavljaju se neovisno o tome je li se kvar zaista dogodio. Često se pri definiranju konstantnog ciklusa uzima onaj u kojem na manje od 1 – 2 % opreme dođe do zastoja uzrokovanog kvarom, što znači da bi većina opreme na kojoj se provodi aktivnost održavanja mogla normalno raditi i dvostruko duži interval od definiranog.

Prednost provođenja takve strategije prvenstveno je u mogućnosti planiranja aktivnosti održavanja i reduciranoj vjerojatnosti pojave kvarova sa ozbiljnim posljedicama [15]. Autori u izvještaju [16] definiraju sljedeće prednosti primjene preventivnog nad korektivnim održavanjem:

- ekonomičnost u troškovno-intenzivnim procesima,
- fleksibilnost prilagodbe periodičnim pregledima,
- povećan životni vijek opreme,
- energetska učinkovita strategija,
- smanjena mogućnost pojave kvarova,
- 12 – 18 % niži troškovi održavanja.

Nedostatak primjene takve strategije, uz to što se i dalje mogu pojaviti nepredviđeni (slučajni) kvarovi, krije se u, često, prevelikom opsegu aktivnosti održavanja i utrošenih doknadnih dijelova, što posljedično donosi smanjenu učinkovitost proizvodnog procesa uvjetovanu nepotrebnim održavanjem.

Strategija planskog preventivnog održavanja odgovarajuća je kod opreme za koju je moguće pouzdano odrediti vrijeme između kvarova, odnosno za onu opremu kod koje je poznata vremenska funkcija degradacije komponenti.

Za opremu, kod koje postoji statistički značajna varijacija oko srednje vrijednosti vremena između kvarova odnosno preostalog vijeka trajanja, preventivno provođenje aktivnosti održavanja najučinkovitije je provoditi primjenom strategije održavanja po stanju.

2.2.3 Održavanje po stanju

Održavanje po stanju, u literaturi često nazivano i prediktivno održavanje (engl. *predictive maintenance*, PdM) moderna je strategija održavanja, koja se temelji na procjeni stvarnog stanja opreme i mogućnosti predikcije pojave kvara. Oprema u radu se kontinuirano ili periodički prati, a aktivnosti održavanja planiraju se u trenutku kad se analizom prikupljenih podataka utvrdi postojanje trenda degradacije funkcije utjecaja na radne sposobnosti, odnosno znakovi nastanka kvara. Takvo načelo strategije održavanja po stanju omogućuje pravovremene aktivnosti održavanja, tj. maksimiziranje funkcijskog vijeka opreme uz nizak tj. prihvatljiv rizik neplaniranog zastoja uzrokovanog kvarom.

Razvoj IoT-a i koncepta Industrije 4.0 omogućuje, kako na globalnoj tako i na lokalnoj razini, međusobno povezivanje sve većeg broja objekata te pristup podacima koji su povezani s tim objektima, npr. podacima sa senzora ili procesnim podacima. U području održavanja, to potencijalno znači porast količine dostupnih podataka iz kojih se u realnom vremenu može procijeniti stanje opreme i planirati aktivnosti održavanja.

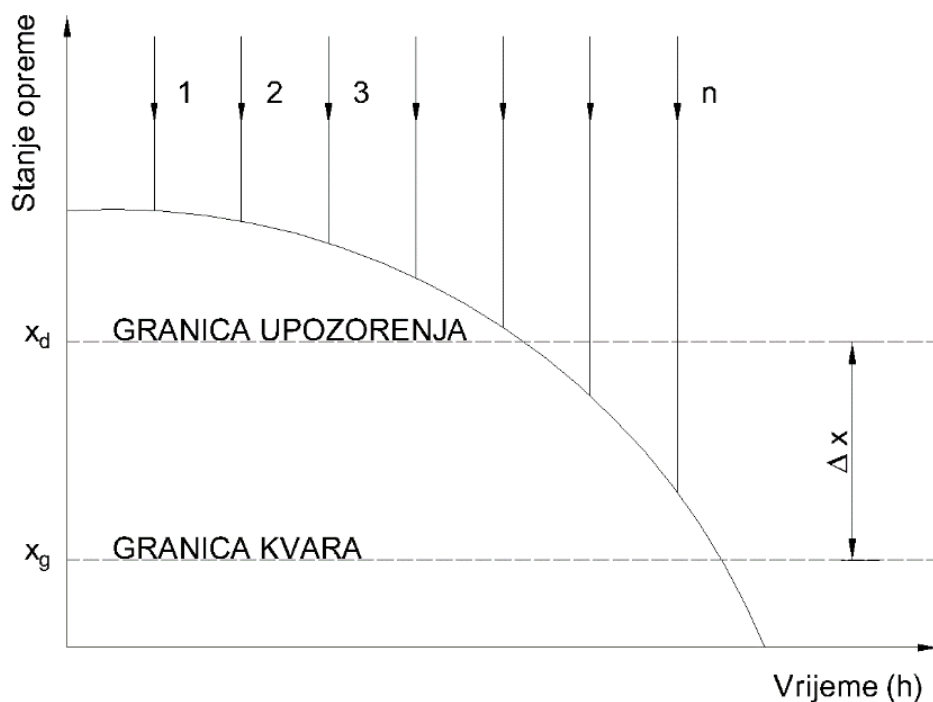
Cilj upotrebe strategije održavanja po stanju je uočavanje mogućih nadolazećih kvarova i otkazivanja kako bi se održavanje provodilo tek kada je to potrebno što rezultira dužim vremenskim periodima eksploatacije opreme, te kraćim i rjeđim zastojima zbog održavanja što se u konačnici odražava i na troškove održavanja. Oprema je ta koja određuje kada će se određene aktivnosti održavanja provesti. To osigurava dovoljno vremena prije otkazivanja opreme ili pada performansi ispod dozvoljene razine, za naručivanje doknadnih dijelova, ako oni nisu na skladištu te da se unaprijed isplanira vrijeme remonta kako bi se što manje utjecalo na proizvodne cikluse i prateće tokove materijala.

Motivacija za primjenu održavanja po stanju jest činjenica da je 99 % kvarova obilježeno prethodnim promjenama u normalnom radu opreme, odnosno indikacijom da će doći do kvara za vrijeme njezine eksploatacije [17]. Može se zaključiti kako je održavanje po stanju strategija koja povećava sposobnost upravljanja *zdravstvenim* stanjem opreme, pri čemu je moguće izbjeći katastrofalne kvarove i sniziti troškove održavanja tijekom ukupnog vremena korištenja opreme.

Kao primjer, navodi se proces određivanja intervala izmjene ulja brodskog ili industrijskog dizel-generatora (preventivno održavanje po konstantnom ciklusu). Proizvođač te intervale (koji mogu biti vezani uz broj radnih sati, npr. svakih 500 rh, 700 rh, 1 000 rh i sl. ili vezani uz vremensko razdoblje od zadnje izmjene u intervalu od 6 mjeseci, 9 mjeseci ili 12 mjeseci)

temelji na višegodišnjim iscrpnim istraživanjima i iskustvu. Treba imati na umu da se termin izmjene ulja ne odnosi na stvarno trenutno stanje ulja već je presjek velikog broja eksperimenata. Primjenom održavanja po stanju izmjena ulja ne provodi se u određenom intervalu već s obzirom na njegovo stvarno stanje prateći vrstu, veličinu i oblik metalnih čestica, koje se nalaze u ulju kao posljedica trošenja pogonskih dijelova, pojavi vibracija i zvukova koji nisu tipični za rad motora ili promjene viskoznih svojstva koji mogu biti posljedica razlike temperatura i/ili starenja ulja.

Dijagramski prikaz strategije održavanja po stanju prema [18] dan je na slici 2.2.



Slika 2.2 Princip održavanja po stanju [18]

Prilikom kontinuiranog praćenja stanja opreme potrebno je definirati početno, odnosno referentno stanje, a slika prikazuje funkciju ovisnosti stanja opreme o vremenskom razdoblju gdje su definirane sljedeće vrijednosti:

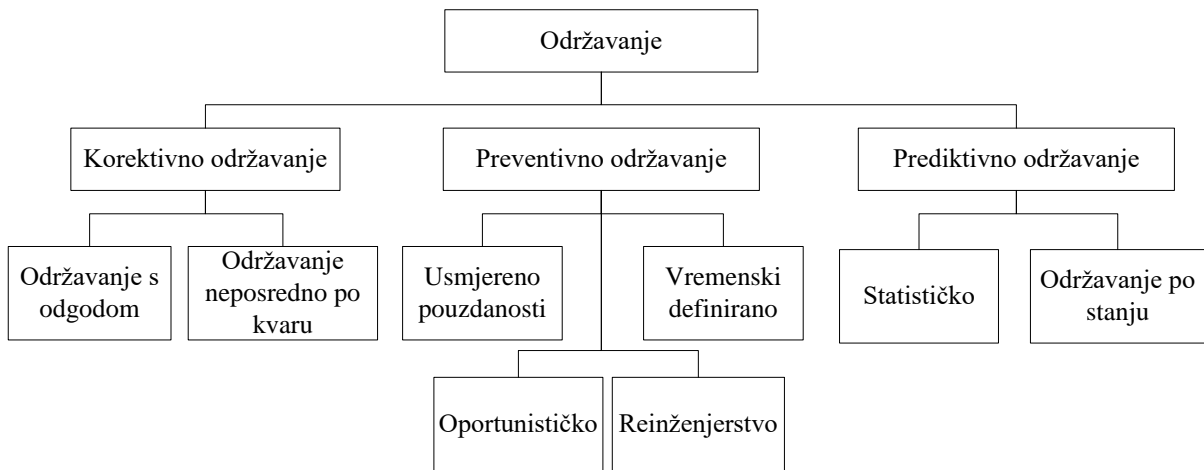
- granica upozorenja (x_d) i
- granica kvara (x_g)

koje se dobivaju numeričkom obradom velikog niza eksperimenata i često su sadržani u pratećoj dokumentaciji opreme. Granica upozorenja predstavlja dopuštenu vrijednost parametara stanja (x_d), a određuje se mjerenjem parametara sustava. Granice upozorenja i kvara definiraju tolerancijsko polje (Δx) koje određuje koliko je dostupno vremena od pojave

relevantnih simptoma do pojave prvih kvarova u skladu sa zakonom promjene stanja promatrane opreme.

Prema strategiji održavanja po stanju, ono počinje kada se prijeđe dopuštena vrijednost parametara stanja (x_d). Vremenski interval (Δt) mora biti dovoljan da bi se provele aktivnosti održavanja po stanju kako oprema ne bi došla u područje pojave kvarova.

Daljnji razvoj industrije, pojava novih proizvodnih koncepata i mogućnosti prikupljanja i obrade podataka dovode do novih klasifikacija održavanja. Kako je prikazano na slici 2.3, autor u [19] predlaže osnovnu podjelu održavanja u tri grupe: korektivno, preventivno (PM) i prediktivno održavanje (PdM), koje se dalje dijeli na prediktivno održavanje temeljeno na statističkim značajkama te stanju opreme.



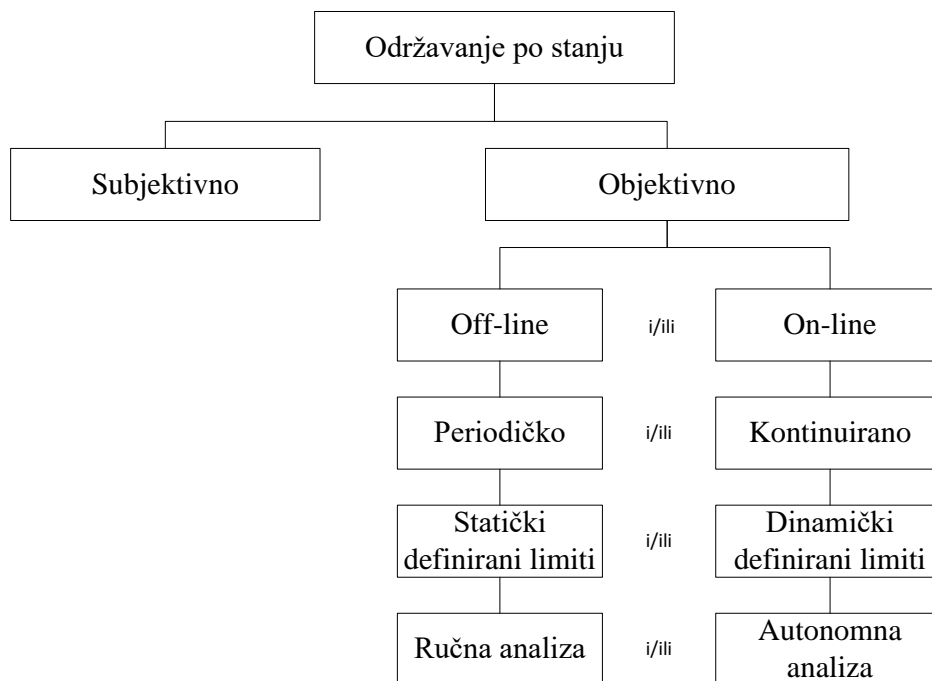
Slika 2.3 Moderna podjela strategija održavanja [19]

Prediktivno održavanje temeljeno na statističkim značajkama pri kreiranju modela za procjenu kvarova uzima u obzir prikupljene podatke o zastojsima opreme uzrokovanim kvarovima, dok modeli održavanja po stanju ovise o podacima kontinuiranog ili periodičkog nadzora stanja opreme. Cilj statističkog pristupa prediktivnom održavanju jest osigurati minimalne ukupne troškove u statistički determiniranom intervalu održavanja, pri čemu nije potrebno korištenje mjerno-akvizicijske opreme za prikupljanje i obradu podataka o trenutnom stanju opreme.

S druge strane, prediktivno održavanje po stanju zahtijeva instalaciju ili korištenje postojećeg sustava za kontinuirano odnosno periodičko mjerenje i prikupljanje podataka, kao što su senzori i sustavi za prikupljanje i obradu podataka. Proces nadzora stanja opreme na način da se mjere i analiziraju podaci prikupljeni sa odgovarajućih senzora smatra se temeljnim procesom prediktivnog održavanja po stanju. Takav proces omogućuje da se aktivnosti održavanja planiraju i izvode samo u trenutku kad su neophodne odnosno prije same pojave kvara koji bi

uzrokovao zastoje. Glavni cilj prediktivnog održavanja po stanju je ostvarivanje mogućnosti donošenja odluka o aktivnostima održavanja na temelju procjene stanja u realnom vremenu, što, posljedično smanjuje broj nepotrebnih aktivnosti održavanja i s time povezane troškove.

Ključna aktivnost prediktivnog održavanja po stanju je proces nadzora stanja, definiran kao aktivnost koja, provedena ručno ili automatski, ima za cilj prikupiti i obraditi podatke i procijeniti stanje opreme [12]. Sam proces nadzora može biti obavljan različitim pristupima i korištenjem različitih razina tehnologije prikupljanja i obrade podataka. Bengtsson u [20] definira pristupe prediktivnom održavanju po stanju, kako je prikazano na slici 2.4.



Slika 2.4 Pristupi prediktivnom održavanju po stanju [20]

Proces nadzora stanja opreme može se provoditi za vrijeme rada opreme (engl. *on-line*), ali i u periodu kad se oprema ne koristi (engl. *off-line*). Nadalje, nadzor se vrši kontinuirano ili periodički. Periodički nadzor najčešće se provodi u određenim intervalima, npr. svakih sat vremena ili na kraju svake smjene, uz pomoć prijenosnih uređaja, kao što su ručni multimetri, analizatori vibracija i buke, ultrazvučni detektori i sl. Također, periodička mjerenja mogu uključivati i evaluaciju temeljenu na ljudskim osjetilima, a koriste se za npr. procjenu stupnja onečišćenja ili utvrđivanje boje odnosno promjene boje lubrikanta. Proces kontinuiranog nadzora uobičajeno uključuje specijalnu mjerno-akvizicijsku opremu. Ograničenje periodičkog nadzora stanja krije se u mogućnosti da se zbog prirode procesa ne prikupe potencijalno važne informacije o stanju opreme između mjerenja. Nasuprot tome, kontinuiran nadzor ne zahtijeva

specijalnu opremu veće vrijednosti, a kontinuirano prikupljanje podataka sa same opreme često znači i veću količinu šuma. U procesu procjene stanja, mogu se koristiti predefimirani (statički) rasponi ili se oni definiraju dinamički tijekom provođenja procesa nadzora. Na kraju, tumačenje rezultata može se temeljiti na procjeni održavatelja ili može biti provedena autonomno od strane inteligentnog sustava za nadzor stanja.

Može se zaključiti, kako su prednosti prediktivnog održavanja nad ostalim strategijama održavanja sljedeće:

- povećanje dostupnosti i pouzdanosti opreme uslijed planiranja i izvođenja aktivnosti održavanja prije potencijalnih zastoja uzrokovanih kvarom;
- smanjenje troškova održavanja kroz smanjenje ukupnog vremena planiranih aktivnosti održavanja, količine neposredno potrebnih doknadnih dijelova te njihovih zaliha; te
- povećanje ukupnog životnog vijeka opreme zbog pravovremenog utjecaja na radnu funkciju i posljedično smanjenog negativnog utjecaja održavatelja uslijed većeg broja aktivnosti održavanja.

S druge strane, mogu se identificirati i ograničenja primjene prediktivnog održavanja:

- potreba za inicijalnim ulaganjem u dijagnostičku opremu i edukaciju rukovatelja;
- kompleksnost opreme i okruženja često uzrokuje teškoće pri utvrđivanju stanja, a također i pri interpretaciji rezultata; te kako
- implementacija prediktivnog održavanja znači i potencijalnu promjenu paradigme održavanja u poduzeću što predstavlja organizacijski izazov.

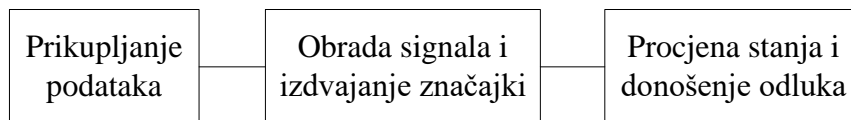
Analogno medicinskom dijagnostičkom procesu, prediktivno održavanje po stanju temelji se na detekciji i dijagnostici stanja upotrebom snimljenih pokazatelja stanja. Kvarovi u ranoj fazi tako mogu biti dijagnosticirani primjenom različitih dijagnostičkih tehnika, a najviše korištene su:

- dijagnostika akustičnom (ultrazvučnom) emisijom,
- dijagnostika analizom sredstva podmazivanja i čestica,
- termografija i
- vibrodijagnostika.

2.3 Tehnike prediktivnog održavanja po stanju

Održavanje po stanju temelji se na mogućnosti snimanja i procjene stanja opreme u pogonu. Takav proces podrazumijeva sposobnost neposrednog snimanja vidljivih, odnosno mjerljivih

značajki (parametara), a koji posredno mogu davati informaciju o stanju opreme, a sve to dok je oprema u pogonu. Jezgru održavanja po stanju čini sustav za mjerenje i pohranu podataka sa određenih tipova senzora ili drugih indikatora, odnosno nadzorni sustav. Nadzorom se podrazumijeva aktivnost, čiji je cilj upravo snimanje i detekcija trenutnog stanja uređaja, koji se promatra, odnosno prikupljanje podataka o kvarovima ali i uzorcima degradacije opreme [21].



Slika 2.5 Proces prediktivnog održavanja po stanju

Generalno, prediktivno održavanje odvija se u koracima (vidi sliku 2.5), pri čemu koraci prikupljanja podataka, obrade signala i izdvajanja značajki ovise o korištenoj tehnici. Prikupljanje podataka je aktivnost pri kojoj se korištenjem odgovarajućeg senzora, snimaju podaci trenutnog stanja opreme. Tako prikupljeni podaci u svom izvornom obliku su najčešće teško upotrebljivi za procjenu stanja. Stoga se u koraku obrade signala i izdvajanja značajki definiraju i izračunavaju značajke, koje u sljedećem koraku omogućuju procjenu stanja i potencijalno donošenje odluka o aktivnostima održavanja.

Postoji veći broj tehnika prediktivnog održavanja po stanju, od kojih su neke postale standardne u primjeni u određenim industrijama. Standardne tehnike uključuju ultrazvučnu analizu, analizu sredstva podmazivanja i čestica, termografiju te vibrodijagnostiku, pri čemu se više različitih tehnika može koristiti za procjenu jednakih tipova kvarova odnosno stanja. Tako na primjer, problem u električnom sustavu otkriven uz pomoć analize ultrazvučnom emisijom, može biti dodatno potvrđen analizom termograma termografske snimke što posljedično donosi i veću sigurnost pri donošenju odluke.

2.3.1 Vibrodijagnostika

Rana dijagnostika kvara ključna je za izbjegavanje većeg kvara i većih gubitaka u ekonomskom smislu. Za procjenu stanja rotacijske opreme, najčešće se koriste vibracijski signali. Izmjereni vibracijski signal uvijek čini kombinacija efekta koji uzrokuje izvor vibracija i efekta koji uzrokuje prijenos vibracija do samog mjesta mjerenja (tj. vibracija strukture opreme).

Čak i u radno ispravnom stanju, oprema generira vibracije. Mnoge takve vibracije izravno su povezane sa periodičkim događanjima u radu uređaja, kao što su rotiranje osovina, zahvati zuba zupčanika, rotiranje električnih polja i dr. Učestalost kojom se takvi događaji ponavljaju često

daje izravnu indikaciju o uzročniku vibracija i zato se mnoge dijagnostičke metode temelje na frekvencijskoj analizi. Neke vibracije se zbog događaja koji nisu u potpunosti sinkronizirani s pogonskim elementima, kao što su izgaranje u cilindru motora (s unutarnjim izgaranjem), ne ponavljaju potpuno periodički sa rotacijom, ali se ponavljaju u nekom zatvorenom ciklusu. Druge vibracije vezane su na tok fluida, često sa nekim unikatnim značajkama kojima se međusobno razlikuju.

Jedan od izazova dijagnostike stanja, koja se provodi pri promjeni vrijednosti mjerenih parametara, jest utvrđivanje je li se ta promjena dogodila na samom izvoru/ima vibracija ili na strukturi opreme, odnosno je li uzrok vibracije promjena stanja opreme ili vanjska uzbuda.

Često se promjena u rezultatima stanja događa zbog promjene na izvoru, kao što su povećanje sile neravnoteže (debalans) ili promjena u spregu sila između zupčanika. S druge strane, promjene u stanju mogu biti rezultat promjene u strukturi opreme, kao što su pukotine u razvoju u kućištu stroja. Također, može se dogoditi i slučaj kad postoje dva ili više uzroka promjene izmjenjenog stanja opreme.

Detekcija kvara rotacijske opreme pomoću nadzora vibracija široko je korištena u industriji jer se najčešći kvarovi, koji su vezani uz nesuosnost ili neravnotežu, mogu detektirati pomoću vibracijskih obrazaca kao odstupanja od standardnih vrijednosti. Vibracije uzrokovane kvarovima pojavljuju se često na specifičnim frekvencijama, čime ih je posljedično moguće povezati sa uzročnom komponentom, dok sama amplituda može indicirati ozbiljnost kvara [22].

Iz gore navedenog, shvatljivo je zašto je analiza vibracija najkorištenija metoda za praćenje stanja rotacijske opreme. Uz to, vibrodijagnostika brzo reagira na promjene i stoga se može upotrijebiti za kontinuirano i periodičko praćenje. Kod, npr. analize ulja, često je potrebno nekoliko dana između prikupljanja uzoraka i njihove analize, iako postoje i neki on-line sustavi za brzu analizu. Također, u odnosu na analizu ulja, analiza vibracija smatra se vjerodostojnijom tehnikom kod procjene kvara na točno određenim komponentama. Naime, analizom ulja mogu se pronaći strane čestice, no često je teško razlučiti o kojoj se komponenti radi, dok analiza vibracija često ukazuje na točno određenu komponentu.

Kriteriji procjene stanja rotacijske opreme na temelju vibracija ovise o širokom rasponu čimbenika, a usvojeni kriteriji mogu se značajno razlikovati za različite vrste opreme. Prema ISO standardu [11], postoje tri primarne veličine pri mjerenju vibracija: pomak, brzina i ubrzanje. Međutim, u većini slučajeva teško je dati apsolutne tolerancije vibracija. Naime, postoji očigledna opasnost lošeg prosuđivanja izmjerenih razina vibracija, opasnih kada nisu,

ili suprotno, neopasnih kada su opasne. Ljudsko iskustvo i tumačenje izmjerenih vrijednosti i dalje ima važnu ulogu u praćenju stanja vibracija.

Nadalje, za procjenu stanja rotacijske opreme, najčešće se koristi snimljeni vibracijski signal prethodno podvrgnut nekoj od metoda obrade signala.

Tehnike obrade signala vibracija, mogu se podijeliti u tri kategorije: obrada u vremenskoj domeni [23], obrada u frekvencijskoj domeni (npr. brza Fourierova pretvorba) i obrada u vremensko-frekvencijskoj domeni (npr. Short-Time Fourier-ova pretvorba, Wigner-ville-ova distribucija i Wavelet pretvorba) [24]. Prema [25], Wavelet pretvorba može biti vrlo korisna u obradi i analizi signala, a time i prepoznavanju uzoraka i dijagnostici kvarova opreme.

Generalizirano, može se reći da za strojeve koji rade na konstantnoj brzini, sinusoidne komponente u rezultirajućim signalima proizlaze iz sinusnih funkcija uzročnih sila na istim frekvencijama, iako će zbog strukturnih nelinearnosti, rezultat obično biti različit od uzroka tj. sadržavati neku razinu harmonika fundamentalne frekvencije, čak i ako je uzročna funkcija relativno *čista*, kao što je npr. slučaj jednostavne neravnoteže.

2.3.2 Dijagnostika akustičnom emisijom

Oprema u radu emitira određeni zvučni obrazac, pri čemu se on razlikuje ovisno o stanju opreme odnosno uvjetima rada. Emitirani zvučni obrazac (engl. *sonic signature*) moguće je snimiti, definirati i prepoznati, što posljedično znači i mogućnost identifikacije promjene zvučnog obrasca kao posljedicu promjene stanja opreme odnosno pojave kvara u ranoj fazi. Zvuk se, za razliku od vibracija koje predstavljaju makroskopsko gibanje strukture, može definirati kao mikroskopska oscilacija u strukturi, odnosno pomak određene strukture na razini molekula. Ultrazvukom se smatraju zvučni valovi frekvencije veće od 20 kHz, a ultrazvučni spektar definiran u području 30 kHz do 2 MHz. Zbog svojih karakteristika pravocrtnog širenja i prolaska kroz pukotine može se koristiti u sljedećim slučajevima:

- otkrivanje propuštanja fluida u cjevovodima,
- dijagnostika stanja ležajeva,
- procjena stanja parnih sustava, odvajača kondenzata i ventila, te
- dijagnostika stanja električnih instalacija.

Za dijagnostiku akustičnom emisijom (engl. *acoustic emission*, AE), ključnu opremu predstavlja akustični senzor. Koriste se piezoelektrični senzori, najčešće tipa PZT (olovo-cirkonijev-titanat), koji se biraju ovisno o interesnoj frekvenciji akustične emisije. Osim samog senzora, opremu čine predpojačalo, pojačalo i sustav za prikupljanje i obradu podataka.

U slučajevima rotacijske opreme, razvijene su brojne tehnike za analizu signala i procjenu stanja ležajeva na temelju analize akustične emisije.

Široko se primjenjuje metoda udarnih impulsa (engl. *shock pulse method*, SPM), tehnika koja procjenjuje stanje ležajeva snimanjem i analizom akustičnog signala u uskom pojasu od približno 36 kHz, pri čemu se snaga ultrazvučnog vala pojačava povećanjem oštećenja [26].

Autori u [27] definiraju tehniku za utvrđivanje izvora oštećenja ležajeva temeljenog na akustičnoj emisiji, dok Al-Dossary i dr. [28] istražuju mogućnosti primjene senzora akustične emisije za određivanje oštećenja radijalno opterećenih ležajeva.

Gowid i dr. [29] uspješno primjenjuju algoritam za selekciju frekvencijskih značajki iz signala senzora akustične emisije te kreiraju model za procjenu stanja sustava centrifugalnog ventilatora.

Iako rezultati istraživanja primjene akustične emisije navode samu tehniku kao učinkovitu u dijagnostici stanja, u praksi se akustična emisija pokazuje vrlo ovisnom o procesnim parametrima što posljedično zahtijeva intenzivno procesiranje signala i pažljivo izdvajanje značajki [30].

Može se zaključiti, kako su potrebna dodatna istraživanja u području primjene akustične emisije u održavanju po stanju, a rezultati ovog rada potencijalno se mogu upotrijebiti kao sljedeća točka istraživanja u smislu automatskog generiranja značajki iz vremenskog signala prikupljenog sa senzora akustične emisije.

Svaka od strategija održavanja industrijske opreme u praktičnoj primjeni ima prednosti i nedostatke, a odgovarajuća strategija održavanja primjenjuje se najčešće na temelju analize kritičnosti opreme. Oprema se po takvom postupku klasificira u različite kategorije kritičnosti, a uvriježena je podjela na kritičnu, važnu, te opremu opće namjene. Generalno, kritičnom se smatra oprema visoke vrijednosti, čiji kvarovi kritično utječu na sigurnost okruženja i kvar kojih zaustavlja proizvodni proces, a ispravan rad štedi energiju. Važna je ona oprema, čiji kvarovi mogu zaustaviti dio procesa ili operaciju, a aktivnosti održavanja zahtijevaju specifična znanja ili su vremenski intenzivna.

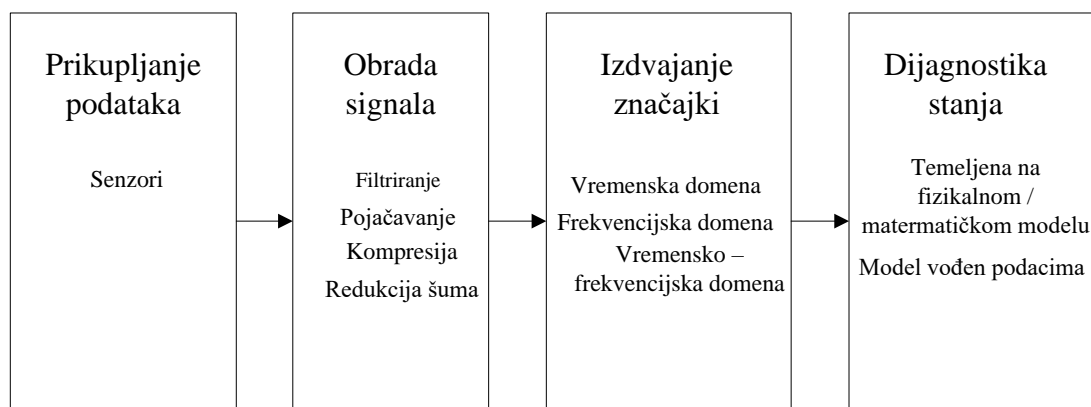
2.4 Inteligentno prediktivno održavanje po stanju

Prediktivno održavanje po stanju trenutno je istraživački najpopularnija i najmodernija strategija održavanja temeljena na mjerenju i analizi parametara s ciljem planiranja aktivnosti održavanja na način da se optimizira životni vijek opreme i bez utjecaja na rizik neplaniranog

zastoja uzrokovanog kvarom [31]. Glavni izazovi prediktivnog održavanja po stanju u proteklih nekoliko desetaka godina bili su vezani uz pristup podacima i informacijama potrebnim za procjenu stanja te sama točnost procjene stanja opreme. Brzi razvoj tehnologija koje omogućuju prikupljanje podataka i napredak u području računalne inteligencije u zadnjih nekoliko godina omogućuje i olakšava razvitak područja inteligentnog prediktivnog održavanja po stanju. Posebno se ističe rastući broj primjera primjene koncepta Industrije 4.0 u kojem se industrijska oprema povezuje na način da može međusobno izmjenjivati podatke i informacije, osiguravajući pritom veliku količinu podataka.

Sama metodologija je izvedena iz fizikalnog fenomena: kada oprema počne mijenjati svoje stanje prema stanju kvara, ljudska osjetila mogu osjetiti promjene na opremi, npr. fluktuacije temperature, pojačane vibracije ili buku. Danas, u doba brzog tehnološkog napretka, senzori su dovoljno napredni da zamijene ljudska osjetila u području održavanja, a da pritom osjetljivost naprednih senzora može biti i nekoliko puta bolja od ljudske. Međutim, sama instalacija senzora nije dovoljna za prediktivno održavanje, budući da je također potrebno izdvojiti informacije iz tih signala, otkriti znanje iz informacija i steći detalje potrebne za procjenu stanja i eventualnu identifikaciju kvarova u ranoj fazi, baš kao što to mora i ljudski mozak. Inteligentno prediktivno održavanje po stanju traži odgovore na pitanja kako učinkovito i prikladno prikupiti podatke, obraditi prikupljene podatke sa senzora, izvući značajke iz obrađenog signala te identificirati stanje i klasificirati kvarove, odnosno predvidjeti potencijalna oštećenja ili degradaciju opreme intenzivnom primjenom računala.

Sljedećim prikazom (vidi sliku 2.6) definiran je proces strategije inteligentnog prediktivnog održavanja po stanju.



Slika 2.6 Proces inteligentnog prediktivnog održavanja po stanju

2.4.1 Prikupljanje, obrada signala i određivanje značajki

Cilj prve faze inteligentnog prediktivnog održavanja po stanju jest efikasno i efektivno prikupljanje podataka, pa se stoga ona temelji na odabiru prikladnih senzora te definiranoj strategiji prikupljanja. Naime, kasnije faze procesa povezane su s vrstom, brojem i lokacijom senzora, kao i njihovim radnim značajkama. Podaci prikupljeni sa senzora vibracija rijetko su korisni u svom sirovom obliku, već se moraju obraditi na odgovarajući način kako bi se izdvojile korisne informacije.

Istražene su i primijenjene brojne učinkovite i korisne metode za obradu signala i izdvajanje značajki u prediktivnom održavanju po stanju. Te se metode uglavnom mogu podijeliti u tri vrste ovisno o domeni izdvajanja značajki (kao što je navedeno u 2.3.1): metode vremenske domene, metode frekvencijske domene i metode vremensko-frekvencijske domene.

Na temelju relevantne literature, u tablici 2.1 navedene su neke od često korištenih metoda za obradu signala i izdvajanje značajki.

Tablica 2.1 Primjena obrade signala i izdvajanja značajki u održavanju

Izvor	Signal	Metoda	Vrsta
[32]	Vibracije	Neuronska mreža	Vremenska domena
[33]	Vibracije	Statističke značajke	Vremenska domena
[34]	Vibracije	EEMD ¹ , PCA ²	Frekvencijska domena
[35]	Vibracije	EEMD	Frekvencijska domena
[36]	Vibracije	FFT ³	Frekvencijska domena
[37]	Vibracije	Kratkotrajna brza Fourierova transformacija	Vremensko- frekvencijska domena
[38]	Vibracije	Metoda omotnice	Vremensko- frekvencijska domena
[39]	Vibracije	Hilbert-Huang-ova transformacija	Vremensko- frekvencijska domena

Cilj prediktivnog održavanja je procjena stanja tehničkog sustava na temelju prikupljenih signala, pri čemu su promjene stanja funkcija vremena. Međutim, u mnogim slučajevima, informacije o stanju pronalaze se u frekvencijskom sadržaju signala, stoga je preporučljivo koristiti i značajke iz frekvencijske domene. Dodatno, uz pomoć vremensko-frekvencijske domene moguće je promotriti značajke iz frekvencijske domene za koje znamo i informacije kad su se dogodile u vremenu.

Proces obrade signala i izdvajanja vremenskih, frekvencijskih ili vremensko-frekvencijskih značajki zahtijeva ekspertno znanje o domeni problema, pri čemu su definirane značajke često primjenjive samo za konkretan slučaj.

Iako sam po sebi zahtjevan i s aspekta konačnog rezultata važan, proces obrade signala i izdvajanja značajki prethodi krucijalnom zadatku dijagnostike, dijagnostici stanja na temelju prikupljenih i obrađenih podataka, odnosno definiranih značajki signala.

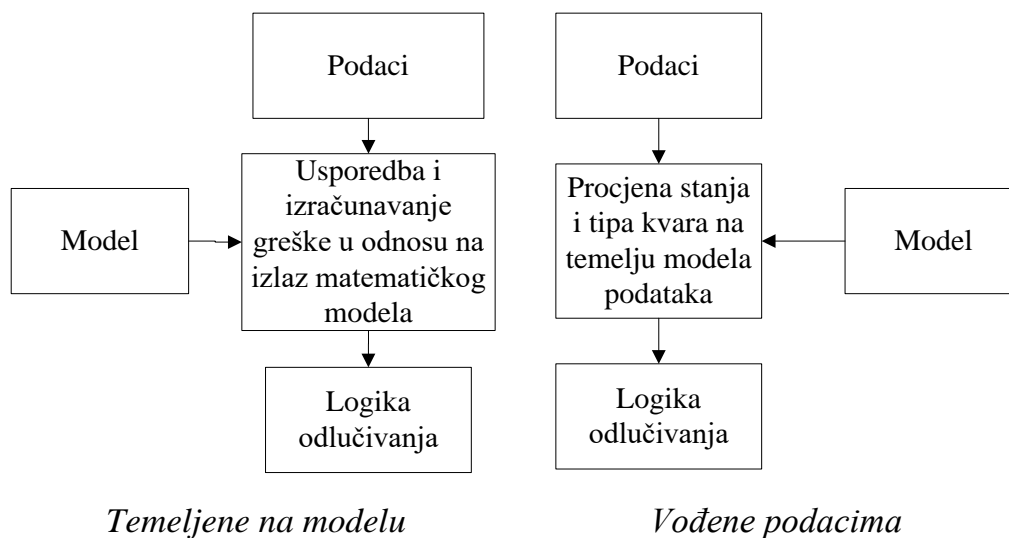
¹ EEMD – empirijsko razlaganje skupa signala (engl. *ensemble empirical mode decompositon*)

² PCA – analiza glavnih sastavnica (engl. *principal components analysis*)

³ FFT – brza Fourierova pretvorba (engl. *fast Fourier transform*)

2.4.2 Inteligentna dijagnostika stanja

Tehnologije koje omogućuju inteligentnu dijagnostiku stanja mogu biti svrstane u dvije glavne kategorije: temeljene na modelu ili vođene podacima (vidi sliku 2.7). Tehnologije temeljene na matematičko-fizičkom modelu rade uz pomoć precizno formuliranog dinamičkog modela sustava i imaju mogućnost detektiranja čak i neočekivanih, nepredviđenih kvarova. Takav sustav radi na temelju usporedbe ulaznih vrijednosti modela i očitavanja senzora (stvarno stanje sustava) te na temelju generiranih razlika donosi odluku o potencijalnom stanju kvara. Nadalje, pristup s fizičkim modelom, pretpostavlja postojanje digitalnog ili realnog blizanca stvarnog sustava, na kojem je moguće simulirati rad opreme, odnosno nužno znači i postojanje takvog sustava te razvoj digitalnog ili realnog blizanca. S druge strane, tehnologije *vođene podacima* (engl. *data-driven*) baziraju svoju logiku odlučivanja na modelu, koji je prethodno treniran na poznatim uzorcima (povijesni podaci) uz pomoć određenog algoritma strojnog učenja te kao takav ima mogućnost donošenja zaključka na temelju prepoznavanja očitavanja senzora kao podskupa već naučenog modela. Za razliku od tehnologije bazirane na modelu, tehnologija vođena podacima nema mogućnost dijagnostike uzroka kvarova, koji nisu anticipirani u modelu, odnosno obuhvaćeni u trening skupu podataka.



Slika 2.7 Tehnologije inteligentne dijagnostike kvarova

Razvoj analitičkih modela sa sposobnošću mapiranja prikupljenih podataka ili definiranih i izvučenih značajki indikatorima, koji mogu točno predvidjeti promjenu stanja komponente ili pojavu kvarova predstavlja složeni inženjerski problem. Posljednjih godina, razvojem tehnika podržanih umjetnom inteligencijom, dokazane su njihove bolje performanse u odnosu na analitičke modele s klasičnim pristupima [40]. Prema pretraženoj literaturi, modeli temeljeni

na umjetnim neuronskim mrežama najviše su istražena i najkorištenija tehnologija umjetne inteligencije za inteligentno održavanje po stanju.

2.5 Umjetne neuronske mreže

U početnim stadijima razvoja umjetne inteligencije 60-tih godina (paralelno razvoju računala) prevladavalo je mišljenje da je samo pitanje vremena kada će ljudski rod biti u mogućnosti proizvesti računalo koje će biti u stanju oponašati rad ljudskog mozga, te da je taj problem jedino povezan sa brzinom procesora računala. Iako je umjetna inteligencija evolvirala, problem je još daleko od rješenja. Razvijene su razne tehnike umjetne inteligencije, koje su prikladne za rješavanje specifičnih problema, a njihov način rješavanja problema može se smatrati *inteligentnim*. Općenito, umjetnom inteligencijom se smatraju automatske računarske procedure bazirane na logičkim ili binarnim operacijama, koje *uče* obavljati neku zadaću na temelju grupe primjera, a naučeno mogu primjenjivati u novoj, nenaučenoj situaciji odnosno s novim podacima. Može se reći da umjetna inteligencija mora oponašati ljudski proces rasuđivanja kako bi rezultat bio donošenje odluka u novim situacijama.

Ideja umjetne neuronske mreže razvijena je iz niza pokušaja modeliranja fiziologije mozga čovjeka, s ciljem razumijevanja i objašnjenja kako isti funkcionira. Ta ideja podrazumijeva kreiranje modela sposobnog obrađivati informacije, analogno aktivnostima mozga čovjeka. Danas je poznata činjenica, kako aktivnost bilo kojeg čvora mozga čovjeka predstavlja zbroj tendencija svih ostalih čvorova da pobude promatrani čvor. Ova temeljna tvrdnja poslužila je za izgradnju osnovne strukture umjetnog neurona. Naime, ako se određeni čvor u mozgu čovjeka zamijeni neuronom, onda se aktivnost neurona može modelirati kao zbroj ulaza neurona sa nekim težinskim faktorom. Takvi ulazi su pomnoženi određenim težinskim faktorima. Prema tome, aktivnost umjetnog neurona ovisi o: broju ulaza (veza) iz okoline (okruženja) neurona, intenzitetu tih veza (iznosu težinskih faktora), te o pragu osjetljivosti koji stanje neurona mora dosegnuti prije nego što pošalje odziv u okolinu neurona, koju čine ostali neuroni umjetne neuronske mreže. Postupak učenja neuronske mreže izvodi se podešavanjem težinskih parametara učenja između slojeva mreže s ciljem da se pogreška na izlazu mreže minimizira, tj. da se vrijednosti izlaza približe željenim.

U istraživanjima provedenim u području održavanja po stanju, korištene su različite vrste umjetnih neuronskih mreža, kao što su unaprijedne mreže sa povratnim prostiranjem pogreške (engl. *back-propagation neural networks*, BPNN), Neuronske mreže sa radijalnim baznim funkcijama (engl. *radial basis function neural networks*, RBFNN), samoorganizirajuće mape (engl. *self-*

organizing maps, SOM), povratne neuronske mreže (engl. *recurrent neural networks*, RNN), dinamičke neuronske mreže (engl. *dynamic neural networks*, DNN), neuro-fuzzy mreže (engl. *neuro-fuzzy networks*, NFN) i dr. Najnovija istraživanja počela su s primjenom dubokih struktura, kao što su naslagani autoenkoderi (engl. *stacked autoencoders*, SAE), Bayesove mreže (engl. *deep belief networks*, DBN) i konvolucijske neuronske mreže (engl. *convolutional neural networks*, CNN).

Tablica 2.2 sadrži popis literature pretražene tijekom istraživanja, odnosno radove koji se odnose na primjenu različitih tehnika umjetne inteligencije u prediktivnom održavanju po stanju.

Tablica 2.2 Literatura – neuronske mreže u prediktivnom održavanju po stanju

Izvor	Područje	Vrsta
[42]	Detekcija kvarova ležajeva	BPNN
[43]	Procjena troškova održavanja	BPNN
[44]	Dijagnostika kvara elektroničkih komponenata projektila	BPNN
[45]	Dijagnostika kvara zupčanika	BPNN
[46]	Dijagnostika kvara ležajeva	BPNN
[47]	Dijagnostika kvara elektroničkih komponenata	RBFNN
[48]	Dijagnostika kvara elektroničkih komponenata	RBFNN
[49]	Dijagnostika kvara sustava napajanja	RBFNN
[50]	Dijagnostika kvara ležajeva	RBFNN, BPNN
[51]	Predikcija kvara zrakoplovnih motora	SOM
[52]	Detekcija kvara klipova zrakoplovnih motora	SOM
[53]	Detekcija anomalija i prediktivno održavanje	SOM
[41]	Detekcija kvara rotora	RNN
[54]	Detekcija stanja rotacijske opreme	DNN / NFN
[55]	Dijagnostika kvara ležajeva	SAE

Izvor	Područje	Vrsta
[56]	Dijagnostika kvara hidrauličke pumpe	SAE
[57]	Dijagnostika kvara indukcijskog motora	SAE
[58]	Detekcija kvarova ležajeva	DBN
[59]	Dijagnostika kvara indukcijskog motora	DBN
[60]	Detekcija kvara CNC strojeva	DBN
[61]	Detekcija kvara kompresora	DBN
[62]	Identifikacija kvara rotacijske opreme	DBN
[63]	Detekcija stanja rotacijske opreme	CNN
[64]	Dijagnostika kvara ležajeva	CNN
[65]	Dijagnostika kvara ležajeva	CNN
[66]	Dijagnostika kvara ležajeva	CNN
[67]	Dijagnostika kvara ležajeva	CNN
[68]	Dijagnostika stanja rotacijske opreme	CNN

Prema literaturi navedenoj u tablici 2.2, može se utvrditi da većina konvencionalnih AI modela kao što su BPNN, RBFNN i SOM, imaju sposobnost otkrivanja trenutnih informacija o greškama. Na taj način moguće je rješavanje problema dijagnostike kada se radi o manjoj količini podataka i kad modeliranju prethodi proces izdvajanja značajki. Međutim, prilikom rada sa velikom količinom sirovih industrijskih podataka, konvencionalni modeli nemaju adekvatnih mogućnosti procjene stanja i dijagnostike kvarova. U narednom odjeljku predstavljene su sustavi za identifikaciju i klasifikaciju kvarova temeljeni na učenju, kao i njihove prednosti u provedbi prediktivnog održavanja koje su nužne u rješavanju trenutnih izazova ove strategije održavanja.

2.6 Duboke neuronske mreže

Duboke neuronske mreže, u literaturi često nazivane i jednostavno *duboko učenje*, mogu se definirati kao konstrukcije dubokih hijerarhijskih modela s mogućnošću strojnog učenja iz podataka. Povijesno gledano, mogu se identificirati ovi trendovi dubokog učenja [69]:

- duboko učenje nije nova tehnologija, već tehnologija za koju su u tijeku razvoja korištena različita imena, kao što su *kibernetika* od 1940. do 60-ih godina prošloga stoljeća ili *konekcionistički model* 80-ih do 90-ih godina prošloga stoljeća. (Na korištene nazive se prvenstveno odražavaju različita filozofska stajališta povezana sa epohom u kojoj se pojedini naziv koristi);
- duboko učenje postaje korisnije s povećanjem količine dostupnih podataka nužnih za učenje;
- modeli dubokog učenja s vremenom postaju složeniji, jer se poboljšala infrastruktura računalnog hardvera i softvera za duboko učenje; te
- duboko učenje rješava sve složenije aplikacije s povećanjem točnosti tijekom vremena.

Jedna od glavnih pretpostavki modela s dubokom arhitekturom jest da je za učenje modela s velikom količinom podataka potrebna hijerarhijska prezentacija tih podataka na nižim, posrednim razinama. Takva definicija sama po sebi određuje razliku dubokog učenja u usporedbi s konvencionalnim (plitkim) učenjem, čijim se prototipom može smatrati napredna neuronska mreža s po jednim ulaznim, skrivenim i izlaznim slojem. Inicijalne metode dubokog učenja razvijene su iz unaprijedne umjetne neuronske mreže s povratnim prostiranjem pogreške. Međutim, istraživači su 80-ih godina počeli shvaćati kako je u praksi iznimno teško primijeniti učenje temeljeno na algoritmu povratnog prostiranja pogreške (engl. *back-propagation algorithm*, BP) za duboke neuronske mreže s više skrivenih slojeva. U Hochreiterovom istraživanju [70], koje se danas smatra prekretnicom eksplicitnog dubokog učenja, formalno je identificiran glavni uzrok te poteškoće: tipične duboke neuronske mreže suočavaju se s problemom nestajućeg ili eksplodirajućeg gradijenta. Sa standardnim aktivacijskim funkcijama, kumulativni signali pogreške u procesu povratnog prostiranja ili propadaju ili eksponencijalno rastu kroz broj slojeva. Buduća istraživanja dubokog učenja temeljila su svoje pretpostavke upravo na prije navedenom saznanju, a kroz godine je razvijeno nekoliko pristupa:

- ublažavanje problema kroz inicijalno učenje bez nadzora za stvaranje hijerarhije (Osnovna ideja je trenirati svaki sloj bez nadzora kako bi predvidio svoj sljedeći ulaz. U ovaj pristup spadaju Bayesove mreže, koje se sastoje od naslaganih ograničenih Boltzmanovih strojeva (engl. *restricted Boltzmann machines*, RBM));
- mreža s dugotrajnom kratkoročnom memorijom (engl. *long term short memory*, LSTM) spada u tip povratnih neuronskih mreža, koje reduciraju fundamentalni problem kroz posebnu arhitekturu sastavljenu od jedinica za izračunavanje pogrešaka kao funkcija

aktivacije i identiteta, a koje otkrivaju važnost događaja koji su se dogodili tisuće diskretnih vremenskih koraka prije;

- napredak u proizvodnji hardvera čini današnja računala, osobito računala s mogućnošću izračunavanja na grafičkoj procesorskoj jedinici (engl. *graphical processing unit*, GPU), mnogo snažnijima od računala iz ranih 90-ih, atakva računalna snaga omogućuje izračun BP algoritma za više slojeva u razumnom roku; te
- prostor matrica težina neuronske mreže također se može pretraživati bez oslanjanja na gradijente grešaka, čime se izbjegava problem potpunog nestajanja gradijenta, što znači kako slučajno nagađanje ponekad djeluje bolje od sofisticiranih metoda [69].

2.7 Primjena dubokog učenja u prediktivnom održavanju po stanju

Primjena tehnika umjetne inteligencije u području održavanja po stanju izazovan je proces s obzirom na sve veću dostupnost podataka s jedne strane te manjak kvalitetnih procedura za obradu podataka i izdvajanje značajki procesa s druge strane. Naime, konvencionalne tehnike strojnog učenja i umjetne inteligencije ograničene su na učenje iz onih signala, kojima su prethodno izdvojene značajke, tj. ne postižu dobre rezultate prilikom učenja iz sirovih signala u onoj formi u kojoj su prikupljeni. Od samih početaka primjene umjetne inteligencije, za formiranje sustava strojnog učenja bila je potrebna i značajna stručnost u postupcima za izdvajanje i odabir značajki, koji transformiraju sirove podatke u prikladnu internu prezentaciju ili vektor značajki iz kojega podsustav učenja, često klasifikator, može otkriti ili klasificirati uzorke ulaznog sloja. Prema pregledu dosadašnjih istraživanja (tablica 2.2), može se utvrditi kako je primjena dubokog učenja započela i u području održavanja po stanju.

Duboko učenje čini niz algoritama koji se mogu primijeniti za modeliranje, aproksimaciju ili mapiranje apstraktnih podataka na visokoj razini. Suština dubokog učenja sastoji se od načina na koji je moguće izračunati hijerarhijske značajke ili prikaze iz objektivnih podataka.

U radu [71] sažeti su relevantni radovi o dubokom učenju i definirane razlike plitkih i dubokih metoda učenja. Kao ključna prednost dubokog učenja navedena je sposobnost samoučenja relevantnih značajki iz samih podataka kroz opći postupak samoučenja.

U novije vrijeme, primjena dubokog učenja najzapaženiju ulogu ima u području računalnog vida i obradi slika [72,73], analizi govora [74–76] te medicinskom inženjerstvu [77–79].

Trenutno se u području prediktivnog održavanja može pronaći nekoliko radova, kojim se započinju istraživanja primjene dubokog učenja za detekciju i dijagnostiku stanja opreme. U nastavku je dan pregled dijela relevantnih istraživanja u ovom području.

Jansens i dr. u [63] analiziraju primjenu dubokog učenja za automatsko učenje značajki. Koriste jedan sloj konvolucijske neuronske mreže za učenje značajki vibracijskog signala na primjeru kvara ležaja te zaključuju kako postoji mogućnost primjene tehnologije konvolucijskih neuronskih mreža za učenje značajki. Ipak, postignuti rezultati ukupne točnosti klasifikacije od 93,61 % dokazuju kako postoji prostor za napredak.

U [62], Shaheryar i dr. predlažu korištenje konvolucijske neuronske mreže u kombinaciji sa autoenkoderskom mrežom i potpuno povezanim klasifikacijskim slojem za identifikaciju stadija oštećenja ležajeva na dostupnom skupu podataka.

Slično Shaheryaru, Zhang i dr. u svojim radovima [64,66] ispituju mogućnosti primjene konvolucijskih neuronskih mreža za identifikaciju i klasifikaciju stadija oštećenja ležajeva na dostupnom skupu podataka. Na testnom skupu podataka uspijevaju dokazati kako je primjenom dubokih struktura moguće klasificirati stadij oštećenja ležajeva, čak i ukoliko se podacima u skupu za testiranje doda šum.

Nadalje, Shaobo Li i dr. [65] predstavljaju novi način korištenja konvolucijskih neuronskih mreža u klasificiranju stadija oštećenja ležajeva kroz senzorsku fuziju. Predlažu obradu sirovih podataka i primjenu korijena srednje vrijednosti kvadrata (engl. *root mean square*, RMS) frekvencijskog spektra signala kao ulaznog signala za učenje.

U [67], Hoang i Kang istražuju primjenu konvolucijske neuronske mreže za modeliranje stadija oštećenja ležajeva na dostupnom skupu podataka. Koriste sirovi vibracijski signal u vremenskoj domeni sa jednog jednoosnog senzora kao ulaz u model, no obrađuju ga na način da pretvaraju 1-D signal senzora u 2-D slike u sivim tonovima, koje dalje koriste za učenje. Napominju kako ne postoji standardna metoda za odabir ispravnih hiperparametara konvolucijskih slojeva, o čemu značajno ovisi proces učenja.

Wen i dr. [68] predlažu korištenje konvolucijske neuronske mreže temeljene na LeNet arhitekturi, a također definiraju postupak konverzije signala u 2D vibracijsku sliku sivih tonova kao način pretprocesiranja. LetNet je jedna od prvih i najjednostavnijih arhitektura korištenih u obradi slika te postaje standardni predložak pri razvoju novih arhitektura. Model mreže testiraju na trima dostupnim skupovima podataka čiji rezultati potvrđuju potencijal za daljnja

istraživanja u tom području. Nadalje, ističu kako je za računalno intenzivan postupak učenja uputno provoditi ga na grafičkom procesoru.

2.8 Identifikacija i klasifikacija kvarova u ranoj fazi temeljena na dubokom učenju

U posljednjih nekoliko desetljeća predloženi su i razvijeni različiti tipovi algoritama za duboko učenje. Duboko učenje obuhvaća veći broj tehnika strojnog učenja, koje obilježava upotreba većeg broja hijerarhijskih slojeva, a s ciljem postizanja veće točnosti i kod rješavanja problema većeg stupnja složenosti.

U ovom odjeljku, dat će se pregled najčešće korištenih arhitektura dubokog učenja, a koje imaju potencijal primjene u prediktivnom održavanju.

Povratne neuronske mreže, za razliku od unaprijednih neuronskih mreža, kao ulazni podatak ne koriste samo informacije koje dolaze određenim slijedom, nego i informacije koje su obrađene nekoliko koraka unatrag. Kombinirajući te informacije, povratne mreže donose odluke na osnovu više prethodno viđenih podataka. Već spomenuti problem nestajućeg ili eksplodirajućeg gradijenta pri korištenju dubokih struktura, u povratnim neuronskim mrežama nastoji se riješiti uvođenjem LSTM jedinica. LSTM jedinice održavaju vrijednosti pogreške konstantnom, tako povratna mreža uči u mnogo više koraka bez da gradijent nestane ili eksplodira. LSTM jedinica se može shvatiti kao ćelija s tri različita stanja ili vrata, a to su *ulaz*, *izlaz* i *brisanje*. Svako od stanja ili vrata ima vlastiti vektor naučenih težina na osnovu kojeg odlučuje treba li ulazni signal dopustiti ulaz u iterativni proces, izlaz iz procesa ili ga treba izbrisati iz procesa. Iako LSTM jedinica rješava problem nestajućeg gradijenta, s druge strane unosi više parametara učenja. Na temelju navedenog, može se utvrditi kako duboke strukture temeljene na ovoj strukturi mogu biti korisne pri obradi vremenskih seriji. U prediktivnom održavanju, Zhao i dr. [80] navode kako većina podataka dolazi sa senzora, koji su po prirodi vremenske serije. U svom radu istražuju mogućnosti primjene LSTM modela za predviđanje istrošenosti alata CNC stroja. Također uspoređuju LSTM s drugim referentnim metodama kao što su linearna regresija, regresija pomoću metode potpornih vektora (engl. *support vector machines regression*, SvMR) i višeslojni perceptron. Rezultati ukazuju kako LSTM ima sposobnost naučiti smislene prikaze iz sirovog signala i generalno daje bolje performanse od uobičajenih metoda s podacima u kojima je prisutna visoka vremenska ovisnost.

Mreža naslaganih autoenkodera ili kraće autoenkoder spada u nenadzirane tehnike učenja. Autoenkoder je algoritam nenadziranog učenja pomoću kojeg je moguće izdvojiti značajke iz ulaznog signala bez dodatnih informacija o klasi ili oznaci tog signala. Svrha autoenkodera je pronalaženje interne reprezentacije uzoraka, odnosno njihovo optimalno kodiranje koje će na izlazu dati uzorak što je moguće sličniji ulazu.

SAE se sastoji od dva dijela, enkodera i dekodera, odnosno ulaznog, jednog ili više skrivenih, i izlaznog sloja. Enkoder vrši kompresiju visoko-dimenzionalnih podataka preslikavajući ulaz na skriveni sloj, a time skriveni sloj uči informacije o ulaznim podacima bez pomoći odnosno označavanja podataka. Naučene informacije spremaju se kao težine skrivenog sloja. Dekoder vrši rekonstrukciju aproksimirajući ulazni signal iz skrivenog sloja, a zadatak cijelog autoenkodera je naučiti skup težina, koji će najtočnije aproksimirati zadani ulaz. Što je ulazni signal složeniji i s više dimenzija, to je potrebno više skrivenih slojeva unutar autoenkodera kako bi se pravilno rekonstruirale mnoge nelinearnosti signala. Sami izlaz autoenkodera nije toliko zanimljiv jer je to samo vjerodostojna kopija ulaza, ali značajke unutar skrivenog sloja koje karakteriziraju ulaz su ono što je bitno kod rezultata autoenkodera. Na taj način se autoenkoderi mogu koristiti kao nenadzirani algoritam prije učenja nekog drugog nadziranog algoritma, u slučaju da ulazni podaci nisu u potpunosti označeni.

SAE se često koristi u smanjivanju dimezionalnosti uzoraka (umjesto metode PCA) ili kao pomoć učenju dubokim klasifikatorima.

Zbog jasnog hijerarhijskog odnosa između dva sloja, SAE ima sposobnost učenja značajki ulaznih signala ili skrivenih informacija o kvarovima iz sirovih ulaznih podataka i podjele istih na više razina. U radu [62] autori koriste hibridni pristup konvolucijske i enkoderske mreže, pri čemu autoenkodersku mrežu koriste za nenadzirano učenje vibracijskih značajki signala.

Algoritam ograničenih Boltzmannovih strojeva je vrsta *plitke* neuronske mreže ili vrsta autoenkodera pogodna za redukciju dimezionalnosti podataka, regresiju, klasifikaciju te učenje značajki. Ograničeni Boltzmannovi strojevi mogu se shvatiti kao dvoslojna neuronska mreža koja se sastoji od jednog vidljivog i jednog skrivenog sloja. Naziv *ograničen* potiče od činjenice da su, za razliku od običnih Boltzmannovih strojeva, ukinute bilo kakve veze neurona unutar istog sloja. RBM algoritam spada u skupinu generativnih modela zbog toga što može naučiti raspodjelu vjerojatnosti ulaznih podataka te potom generirati nove uzorke koji odgovaraju skupu trening podataka. Kod problema nepotpunih ulaznih podataka, RBM algoritam također može na temelju predviđanja raspodjele vjerojatnosti nadomjestiti izgubljene ili nedostajuće ulazne podatke.

Težine i pomaci između dva sloja uče se iteracijama, kako bi izlaz vidljivog sloja bio što bolja aproksimacija ulaznog signala. Parametri u skrivenom sloju mogu biti promatrani kao značajke koje karakteriziraju ulazni signal bez prethodnog znanja o prirodi signala, a potom se na RBM algoritam mogu postaviti algoritmi nadziranog učenja kao što je logistička regresija ili neuronske mreže koje mogu izvršiti klasifikaciju ili regresiju.

Kao samostalni algoritam, RBM se mogu slagati jedan na drugi i tako formirati strukture zvane duboke probabilističke mreže ili duboki Boltzmannovi strojevi (engl. *deep Boltzmann machine*, DBM). Razlika između ta dva algoritma je u njihovom načinu spajanja slojeva. Naime, duboke probabilističke mreže djelomično su usmjerene mreže u kojima su sve veze, osim onih između dva najviša sloja usmjerene prema dolje, dok su DBM potpuno usmjereni algoritmi. Usmjerenost veze podrazumijeva da neuron od kojeg veza kreće utječe na neuron u kojeg ista ulazi, ali ne i obrnuto. Prema tome, algoritam duboke probabilističke mreže može biti efikasniji i lakši za naučiti od dubokog Boltzmannovog stroja. RBM-ovi se mogu naslagati jedan na drugi povezujući skriveni sloj jednog RBM-a s vidljivim slojem sljedećeg. Pregled dostupne literature ukazuje kako postoji mogućnost primjene ovakvog tipa mreža u održavanju po stanju. U radu [59] Shao i dr. ispituju mogućnost primjene DBN modela za izdvajanje značajki signala indukcijskog motora te utvrđuju potencijal primjene takve arhitekture u održavanju po stanju. Navode kako je cjelokupno područje još u povojima razvoja te su potrebna daljnja intenzivna istraživanja kako bi se utvrdio stvaran kapacitet primjene.

Konvolucijske neuronske mreže su vrsta neuronskih mreža koja je svojom internom strukturom prilagođena za relativno brzo i efikasno rješavanje problema više-dimenzionalnih ulaza odnosno ulaza koji posjeduju mnoštvo značajki. Svoju glavnu primjenu CNN su stekle u obradi, klasifikaciji i prepoznavanju objekata na slikama, no od nedavno se mogu pronaći i primjeri istraživanja primjene takve arhitekture u području održavanja [62,65–68]. S obzirom na to da su se u ovom radu konvolucijske neuronske mreže koristile kao arhitektura za modeliranje, sam način rada algoritma bit će detaljnije objašnjen u poglavlju 3.

2.9 Sažetak poglavlja

U ovom poglavlju opisan je razvoj prediktivnog održavanja te su definirane i prikazane tehnike i modeli, koje je moguće koristiti za implementaciju navedene strategije. Nadalje, dan je literaturni pregled trenutno aktualnih arhitektura i tehnika dubokog učenja u području inteligentne dijagnostike stanja i identifikacije uzroka kvarova, a za intenzivnije korištene tehnike opisani su korišteni modeli.

Zbog sve veće i sve brže rastuće količine prikupljenih podataka, postoji sve veća potreba za računalnim modelima održavanja po stanju temeljenima na podacima. Takvi modeli mogu rezultirati bržim otkrivanjem znanja, odnosno rasteretiti eksperta/donositelja odluka intenzivnog rada u samom procesu pripreme za donošenje odluka. Područje dijagnostike stanja usmjereno je prema inteligentnoj dijagnostici, a raste primjena dubokih struktura.

Konvolucijske neuronske mreže, iako svoju primarnu ulogu nalaze u obradi slika, posjeduju pogodnosti iskoristive i u području istraživanja ove disertacije te su odabrane kao tehnika za razvoj modela omogućujući automatsko generiranje značajki signala vibracija i klasifikaciju stanja rotacijske opreme.

3. KONVOLUCIJSKE NEURONSKE MREŽE

Konvolucijska neuronska mreža je varijanta višeslojnih unaprijednih neuronskih mreža, a dizajnirana je prvenstveno za obradu slika ili senzornih podataka u obliku matrica, uzimajući u obzir lokalna i globalna stacionarna svojstva [81]. Slično strukturi klasičnog višeslojnog perceptrona odnosno klasične unaprijedne mreže. CNN je mreža sastavljena od *slojeva*, u kojoj je izlaz prethodnog sloja povezan sa ulazom sljedećeg pomoću skupa parametara (težina i pomaka (engl. *bias*)), koji se mogu učiti.

Glavna je razlika u odnosu na višeslojni perceptron u tome što je svaki sloj predstavljen kao skup ulaznih i izlaznih mapa značajki, izračunatih postupkom konvolucije, a u cilju obuhvaćanja različitih perspektiva ulaznih podataka.

Kako je već navedeno, svoju glavnu primjenu CNN su stekle u obradi, klasifikaciji i prepoznavanju objekata na slikama, no od nedavno se mogu pronaći i primjeri istraživanja primjene takve arhitekture u području održavanja [62,65–68].

Generalno, rad konvolucijske mreže može se podijeliti u tri glavne operacije:

- konvolucija,
- aktivacijska funkcija i
- sažimanje.

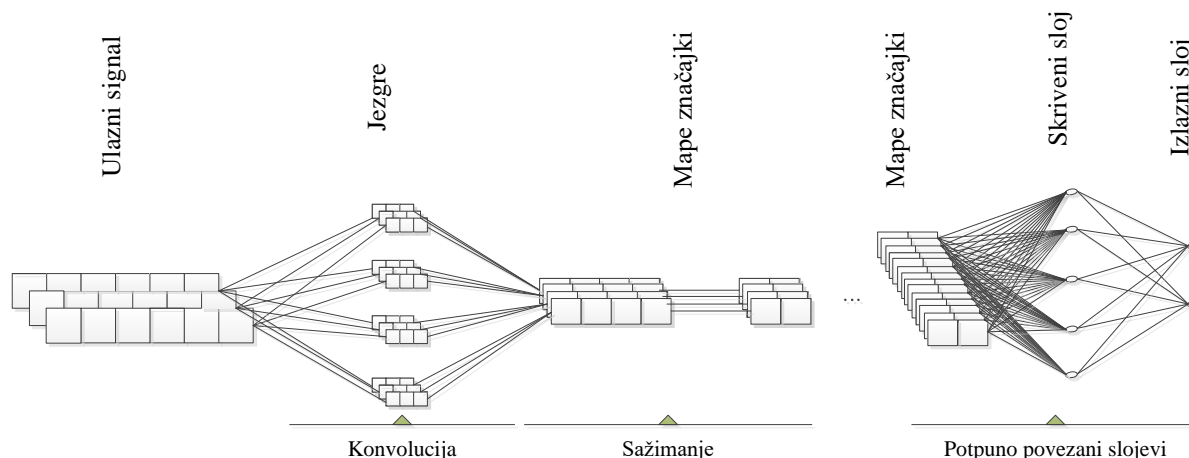
nakon kojih slijedi jedan ili više potpuno povezanih slojeva s ciljem klasifikacije značajki naučenih u konvolucijskim slojevima, kako je prikazano na slici 3.1.

Slojevi se naizmjenice slažu unutar okvira konvolucijske mreže, sve do željene dubine odnosno do ostvarivanja mogućnosti izdvajanja značajki, pomoću kojih se provodi klasifikacija u potpuno povezanom sloju.

3.1 Konvolucijski sloj i aktivacijska funkcija

Konvolucijski slojevi jesu temeljni arhitekturni blokovi konvolucijske neuronske mreže s ciljem izračunavanja operacije konvolucije ulaznog signala i vrijednosti jezgri. Konvolucijski sloj definiran je hiperparametrima: veličinom jezgre, brojem jezgri i korakom.

Svaki konvolucijski sloj može sadržavati više mapa značajki, dok se težine pojedinog konvolucijskog sloja dijele između mapa. Ovakva arhitektura omogućuje učenje mreže s različitim vrijednostima značajki uz zadržavanje vrijednosti parametara.

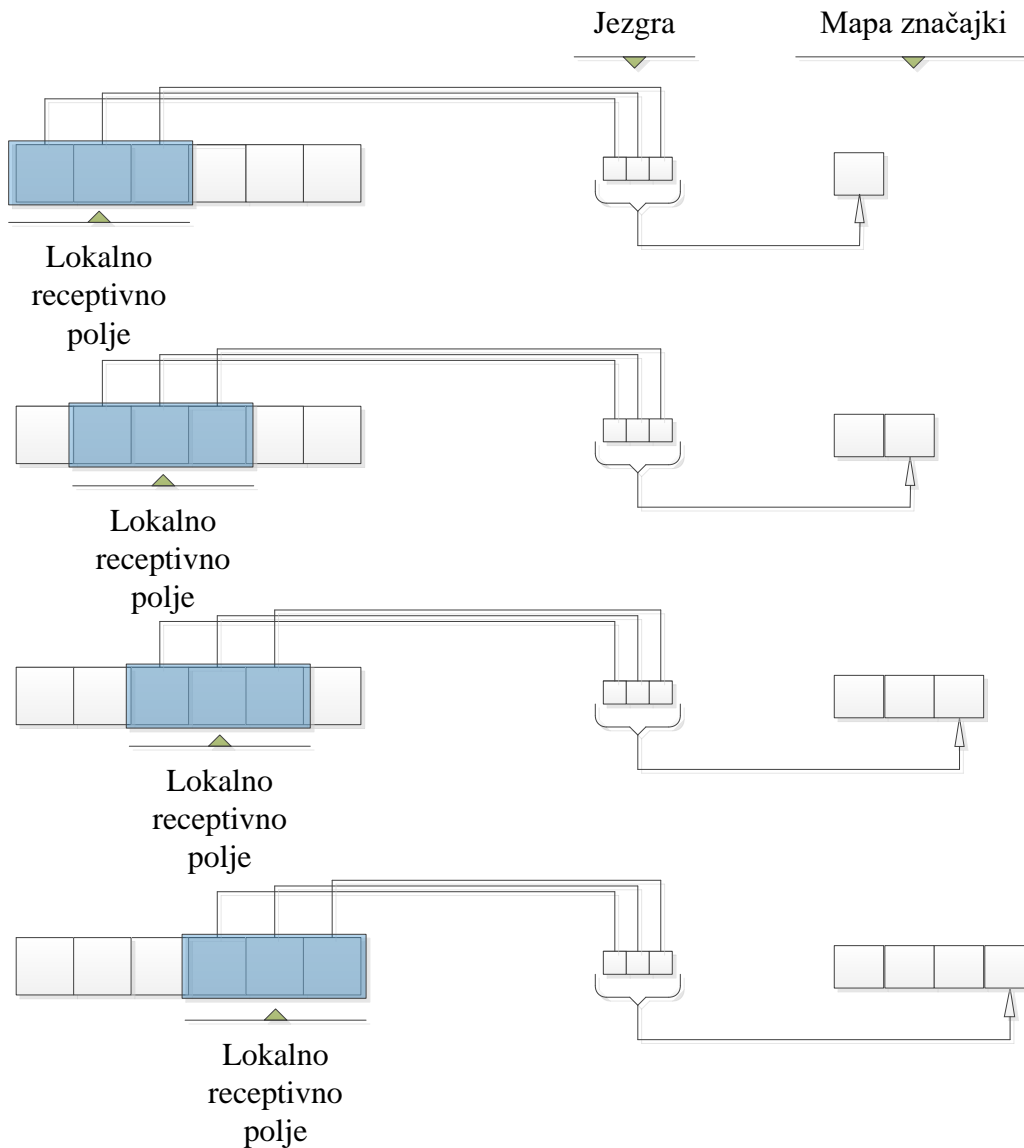


Slika 3.1 Opća struktura konvolucijskih neuronskih mreža

Ulaz u mrežu predstavljen je u ovom slučaju vektorom čija je duljina ujedno i veličina ulaznog sloja. Ulaz u mrežu može biti i matrica (slika) ili polje vrijednosti. Ulazni podaci ulaznog sloja spajaju se, kao i kod klasičnih plitkih struktura neuronskih mreža sa prvim sljedećim slojem. No, za razliku od potpunog povezivanja neurona ulaznog sloja u slučaju klasične unaprijedne neuronske mreže, konvolucijska neuronska mreža spaja svaki neuron konvolucijskog sloja samo s regijom povezivanja (manjom grupom neurona) ulaznog sloja. Regija povezivanja naziva se lokalna receptivna regija ili polje (engl. *local receptive field*), a njezina veličina određena je veličinom jezgre (engl. *kernel*). Nad lokalnim receptivnim poljem odvija se operacija konvolucije, pri čemu se svaki član lokalnog receptivnog polja množi sa prikladnim članom jezgre, a potom se produkti zbrajaju u definirano polje u mapi značajki. Konvolucija nad ulazom u konvolucijskim sloj provodi se množenjem odgovarajućih ulaza sa težinama definiranim u jezgri, njihovim sumiranjem te dodavanjem praga. Potrebno je naglasiti da je pri izračunu vrijednosti za jedno polje potrebno sumirati umnoške ulaza i težina svih dubina. Za izračun vrijednosti sljedećeg polja potrebno je pomaknuti jezgru za definirani pomak i ponoviti postupak.

Dakle, polje u mapi značajki izračunano konvolucijom pritom sadrži informaciju, kojom se predstavljaju vrijednosti svih elemenata receptivnog polja te značajke iz prethodnog sloja. Slika 3.2 ilustrira takvu proceduru. Nakon što se izračunaju sve vrijednosti, kao i kod jednostavne neuronske mreže, dobivene vrijednosti potrebno je provesti kroz aktivacijsku funkciju. Ovakav primjer konvolucije pretpostavlja relaciju između jednog ulaznog sloja odnosno jedne ulazne mape značajki sa jednom izlaznom mapom. Pritom, ulaznom mapom značajki smatra se izlaz iz prethodnog konvolucijskog sloja, osim ukoliko se radi o prvom konvolucijskom sloju koji se

veže na ulaz predstavljen isključivo jednim ulaznim vektorom, matricom odnosno poljem. Svaka veza sloja sa prethodnim slojem predstavljena je jezgrom. Shodno tome, svaka mapa značajki mora se sastojati od onoliko jezgara koliko je mapa u prethodnom, s njom povezanom sloju.



Slika 3.2 Stvaranje mape značajki konvolucijom

Vrijednost mape k' sloja l na lokaciji (i, j) može se izračunati prema jednadžbi (3.1)

$$x_{k'}^l(i, j) = f\left(\sum_{k \in M^{l-1}} \sum_{\substack{0 < m < K_w^k \\ 0 < n < K_h^k}} x_k^{l-1}(i+m, j+n) w_{k'}^l(m, n) + b_{k'}^l\right), \quad (3.1)$$

gdje je k jedna mapa značajki trenutnog sloja $M^l (k \in M^l)$, a $x_k^l(i, j)$ je element na poziciji (i, j) unutar te mape. Veličina jezgre opisana je sa $K_w^k \cdot K_h^k$, a svaka jezgra mape k sloja l nosi vrijednost $w_k^l(x, y)$ na lokaciji (x, y) . Vrijednost b_k^l predstavlja pomak.

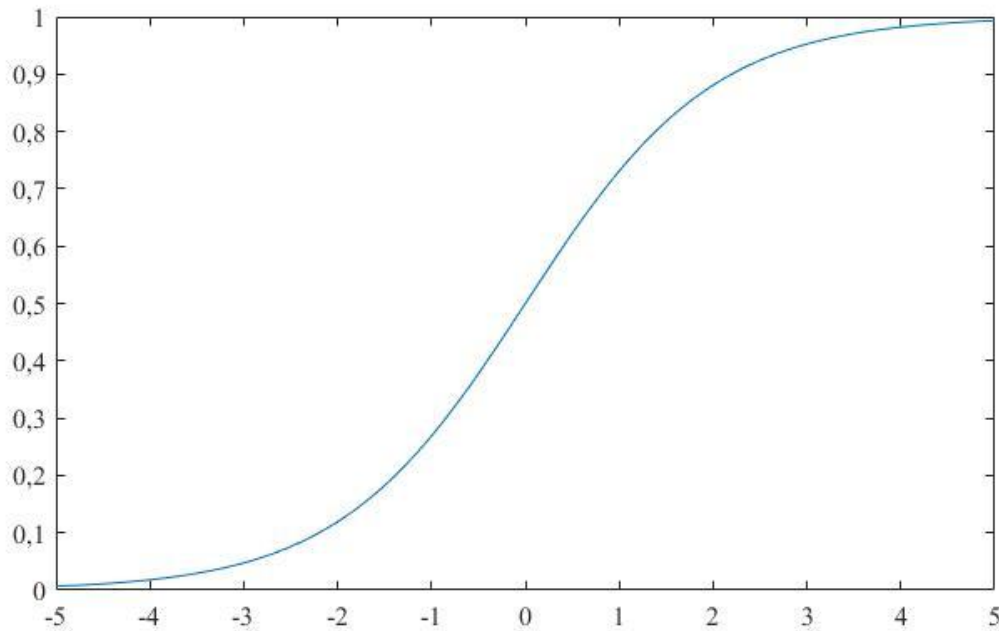
Nakon izračuna vrijednosti za element (i, j) izlazne mape značajki, definiranim korakom (engl. *stride*) prelazi se na sljedeći po redu element $(i + S_w, j)$ ili $(i, j + S_h)$ ukoliko se prelazi u novi redak matrice. Može se koristiti zadani korak $S_w = S_h = 1$, ali i uvećan za određeni faktor, što je isto tako definirano kao hiperparametar pri razvoju modela.

Postupak povezivanja svake mape značajki trenutnog sloja sa svim mapama značajki prethodnog sloja opisuje se kao proces potpunog povezivanja konvolucijskih slojeva. Osim takvog povezivanja, moguće je koristiti i raspršeno (engl. *sparse*) povezivanje, pri kojem nisu sve mape značajki trenutnog sloja povezane sa svim ostalim mapama značajki prethodnog sloja. Posljedično, potrebno je definirati matricu povezanosti dimenzija $M^l \times M^{l-1}$, kojom se utvrđuje koje su mape značajki trenutnog sloja povezane s mapama značajki prethodnog sloja.

Nakon postupka učenja, vrijednosti težina i pomaka, koji povezuju polja ulaznih neurona i skrivene neurone, imati će iste vrijednosti, tj. čitav skriveni sloj dijelit će iste vrijednosti naučenih parametara odnosno prepoznavati istu značajku, na svim mogućim lokacijama izvorne datoteke. Ova karakteristika konvolucijskih neuronskih mreža u literaturi se definira kao dijeljenje parametara (engl. *weight sharing*).

Nakon postupka konvolucije ulaznog signala uz pomoć jezgri, dobivene mape značajki potrebno je, analogno postupku kod potpuno povezanih neuronskih mreža, provesti kroz nelinearne aktivacijske funkcije kako bi se modelirale nelinearnosti. U aktivacijskim funkcijama potrebna je nelinearnost jer je njezin cilj proizvesti nelinearnu granicu odluke putem nelinearnih kombinacija težina i ulaza. Na taj način, moguće je naučiti model da daje nelinearan odziv s obzirom na kombinaciju ulaznih vrijednosti. Za probleme linearne regresije, moguće je također koristiti linearne aktivacijske funkcije u skrivenim slojevima. Ipak, s obzirom na činjenicu kako će se takva mreža neovisno o broju skrivenih slojeva ponašati kao jednoslojna mreža, primjena linearnih aktivacijskih funkcija nije uobičajena. Jedna od najkorištenijih aktivacijskih funkcija jest sigmoidalna funkcija, koja uz svojstvo nelinearnosti i monotonog rasta posjeduje i svojstvo derivabilnosti nužno za učenje pomoću algoritma povratnog prostiranja pogreške. Jednadžba (3.2) i slika 3.3 definiraju i prikazuju navedenu aktivacijsku funkciju.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.2)$$

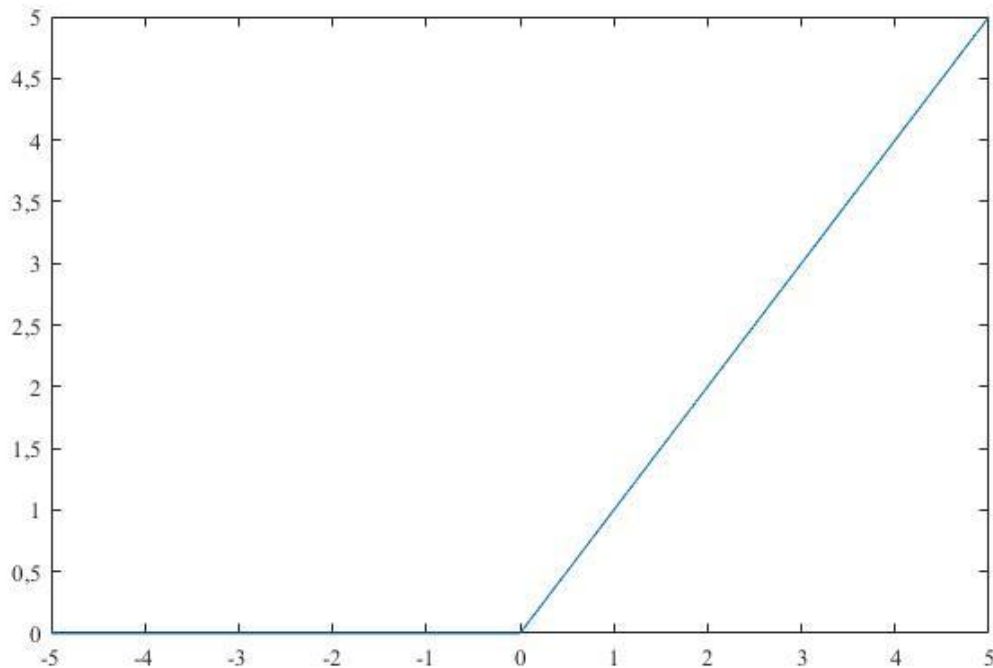


Slika 3.3 Sigmoidalna aktivacijska funkcija

U početnom razdoblju uporabe neuronskih mreža (nakon pojave algoritma povratnog prostiranja pogreške), sigmoidalne funkcije bile su najčešće korištene aktivacijske funkcije neurona. No, s pojavom dubokih arhitektura, njihov intenzitet korištenja se smanjio zbog negativnog efekta nestajućih gradijenata (vidi odjeljak 2.5). Naime, s obzirom da sigmoidalna funkcija ograničava izlaz neurona na interval $(0,1)$, pri čemu se asimptotski teži vrijednosti 1 odnosno nula, neuroni u dubljim slojevima mreže mogu poprimiti male odnosno velike vrijednosti što posljedično znači da i aktivacija neurona (vrijednost zbroja ulaza uvećana za pomak i težine) postane mala ili velika vrijednost. U takvom slučaju, gradijent u toj regiji se smanjuje, a konačna posljedica jest prekid učenja ili vrlo spora konvergencija algoritma učenja uslijed malih gradijenata, koji utječu na promjenu težina. Iako je navedeno moguće rješavati kroz uvođenje dodatnog sloja za normalizaciju podataka (engl. *batch normalization layer*), najčešće se pribjegava korištenju drugih (bržih) aktivacijskih funkcija.

U dubokim arhitekturama, pa tako i u konvolucijskim neuronskim mrežama, uobičajeno se koristi zglobna, tj. ReLU (engl. *rectifier linear unit*) aktivacijska funkcija, definirana jednačinom (3.3) i prikazana na slici 3.4.

$$f(x) = \max(0, x) \quad (3.3)$$



Slika 3.4 ReLU aktivacijska funkcija

Sigmoidalna funkcija u povratnom prostiranju omogućuje množenje gradijenata sa vrijednostima manjim od 1. Na taj način, u dubljim slojevima gradijenti postaju sve manji, a konvergencija sve sporija. Korištenjem nelinearne ReLU funkcije osigurava se konstantna vrijednost gradijenta (gradijent je uvijek 0 ili 1), s obzirom da je rezultat pozitivnog ulaza uvijek različit od nule. Tako definirana aktivacijska funkcija osigurava da prilikom povratnog prostiranja nema gradijenata koji se progresivno smanjuju i postaju sve manji u dubljim slojevima, što za posljedicu ima brže učenje. ReLU aktivacija, zbog prethodno definiranog svojstva pogodna je za korištenje u dubokim arhitekturama. Naime, konstantna vrijednost gradijenta, osim što smanjuje mogućnost nestajućeg gradijenta, osigurava brže učenje. Ipak, negativna vrijednost aktivacije neurona nakon derivacije iznosi nula, što u praktičnom slučaju znači gašenje neurona mreže. Korištenje dodatnog sloja za normalizaciju podataka (engl. batch normalization layer) smanjuje vjerojatnost pojave većeg broja gašenja neurona. Osim navedenog, moguće je koristiti i prilagođenu verziju ReLU aktivacijske funkcije (engl. leaky ReLU), koja i za ulazne vrijednosti manje od 0 dodjeljuje malu pozitivnu vrijednost (npr. $y = 0,01|x|$ za $x < 0$). S obzirom na činjenicu kako je ReLU najkorištenija aktivacijska funkcija u dubokim strukturama, koristit će se u ovom istraživanju.

3.2 Sloj sažimanja

Sažimanje (engl. *pooling*) često je korišten postupak unutar konvolucijskih neuronskih mreža, a provodi se s ciljem smanjenja dimenzija ulaza i povećanja neosjetljivosti mreže kroz uzorkovanje, pri čemu se ne mijenja broj mapa značajki, već samo njihove dimenzije.

Sama operacija sastoji se od grupiranja ulaza na podregije zadane veličine $S_w \times S_h$ uz definirano preklapanje te njihovu prezentaciju jednom vrijednošću.

Prve implementacije [82] koristile su vrijednost aritmetičke sredine, no kasnije su u nekoliko radova [83,84] opisane prednosti korištenja maksimalne vrijednosti regije sažimanja?.

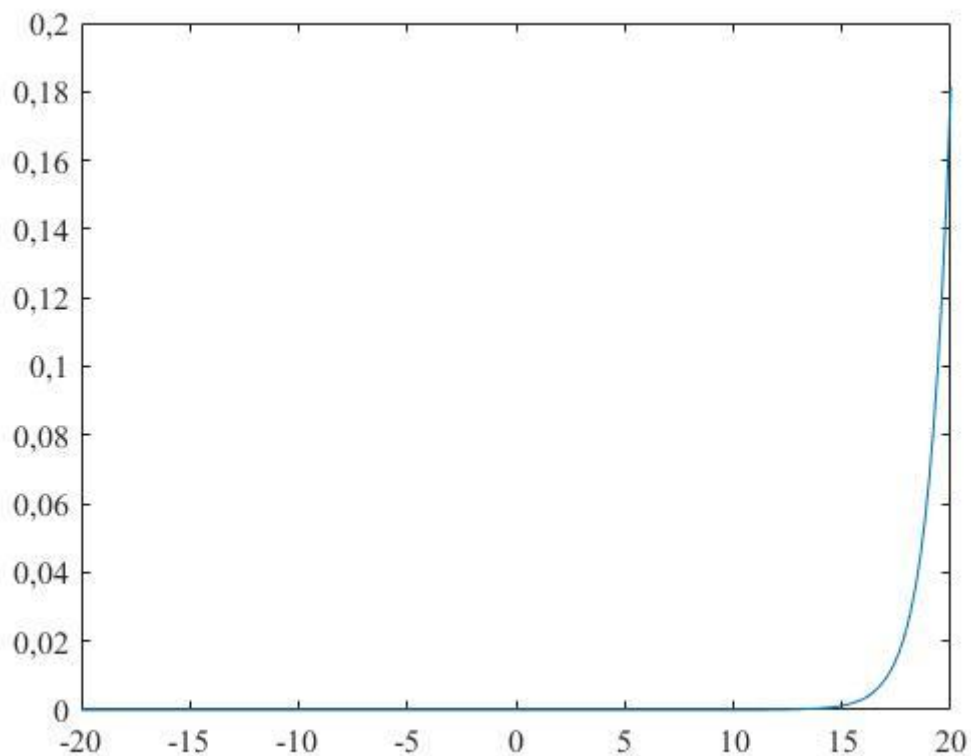
Posljedica sažimanja je redukcija dimenzionalnosti ulaza, što sa sobom povlači smanjenje broja parametara za učenje. Dodatno, sažimanjem se smanjuje osjetljivost konvolucijske mreže na različitost ulaza, što je naročito korisno u slučaju promjena položaja značajke koja se nastoji detektirati jezgrom konvolucijskog sloja unutar promatranog lokalnog receptivnog polja.

3.3 Izlazni sloj

Softmax funkcija jedna je od najpoznatijih funkcija u znanosti i inženjerstvu i intenzivno se koristi u području umjetne inteligencije za mapiranje vektora značajki klasa višeklasne logističke regresije u gustoću razdiobe vjerojatnosti pojedinih klasa.

Sam naziv *softmax* funkcije skraćenica je od složenice *soft argmax*, a matematički predstavlja generalizaciju sigmoidalne funkcije (najčešće korištene kod binarne klasifikacije) sa sposobnošću transformacije k -dimenzionalnog vektora x sastavljenog od neskaliranih realnih vrijednosti u k -dimenzionalni vektor $f(x)$ čije komponente sadrže vrijednosti u intervalu $(0,1)$ te im je zbroj jednak 1, kako je prikazano u jednadžbi (3.4) te na slici 3.5:

$$f(x)_j = \frac{e^{x_j}}{\sum_{i=1}^k e^{x_i}}, j \in \{1, 2, \dots, k\}, \quad (3.4)$$
$$\sum_{j=1}^k f(x)_j = 1.$$



Slika 3.5 Softmax funkcija

Izlaz konvolucijske neuronske mreže u konačnici predstavlja stupanj vjerojatnosti (sigurnosti) mreže u donesenu odluku, pri čemu veći broj označava veću vjerojatnost točne odluke. Softmax funkcija unutar konvolucijske neuronske mreže prethodi klasifikacijskom sloju, a osigurava mogućnost probabilističke interpretacije rezultata odnosno sigurnosti mreže u izračunatu odluku.

3.4 Učenje konvolucijske neuronske mreže

Učenje neuronske mreže predstavlja iterativni postupak određivanja vrijednosti težina neurona u mreži. Kako bi se vrijednosti težina mogle prilagođavati, unaprijed se određuje funkcija gubitka, koja se nastoji minimizirati. Najčešće korišteni algoritam za modifikaciju težina neurona jest algoritam povratnog prostiranja pogreške (engl. *backpropagation*), koji određuje način izračuna gradijenata uz pomoć kojih se ažuriraju težine neurona nakon prolaska podataka kroz mrežu.

Algoritam povratnog prostiranja pogreške moguće je uz modifikacije u konvolucijskim i slojevima sažimanja primjeniti i na konvolucijske neuronske mreže. U sljedećim odjeljcima,

opisan je rad i specifičnosti algoritma povratnog prostiranja pogreške u konvolucijskoj neuronskoj mreži.

3.4.1 Gradijentni spust

Gradijentni spust optimizacijski je algoritam za pronalaženje minimuma derivabilne funkcije. Minimum funkcije izračunava se tako, da se za točku konveksne funkcije izračunava smjer gradijenta (parcijalna derivacija) te se u sljedećem koraku kreće u smjeru maksimuma negativnog gradijenta. Učenje neuronskih mreža podrazumijeva optimiranje funkcije gubitaka E , na način da se uz pomoć algoritma gradijentnog spusta podešavaju težine neuronske mreže. Može se zaključiti kako je potrebno pronaći vektor težina w koji će minimizirati funkciju cilja, odnosno funkciju gubitaka $E(w)$. Najprije je potrebno napraviti malu promjenu iznosa težine w na sljedeći način:

$$w = w + \delta w. \quad (3.5)$$

Promjenom iznosa težina mijenja se i funkcija gubitaka E :

$$\delta E = w^T \nabla E(w), \quad (3.6)$$

gdje je $\nabla E(w)$ gradijentni vektor funkcije gubitaka koji predstavlja smjer u kojem se ostvaruje najveći pad funkcije gubitaka. Najmanja vrijednost funkcije gubitaka ostvarit će se na mjestu gdje je:

$$\nabla E(w) = 0. \quad (3.7)$$

Za minimiziranje funkcije gubitka primjenjuju se iterativni postupci. Uglavnom, iterativne tehnike provode se na način da se izabire nasumična vrijednost $w^{(0)}$ za vektor težina koja se zatim mijenja na sljedeći način:

$$w^{(\tau+1)} = w^{(\tau)} - \Delta w^{(\tau)}, \quad (3.8)$$

gdje τ označuje trenutni iteracijski korak.

Najprimjereniji pristup optimiranju parametara je upravo pomicanje u malim koracima i to u smjeru suprotnom od smjera vektora gradijenta funkcije gubitaka. Ovo se može zapisati u obliku jednadžbe (3.9):

$$w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E(w^{(\tau)}), \quad (3.9)$$

gdje $0 < \eta < 1$ predstavlja koeficijent učenja. Koeficijent učenja (koristi se i izraz *stopa učenja*) jedan je od glavnih hiperparametara s najvećim utjecajem na promjenu rezultata neuronske mreže, a kojeg treba samostalno izabrati kako bi mreža postigla dobru konvergenciju prema najboljem vektoru težina.

Nakon svakog ažuriranja, ponovno se izračunava gradijent za novi vektor težine, te se postupak ponavlja.

Postoje dvije glavne varijante gradijentnog spusta, ovisne o trenutku ažuriranja težina:

– *grupni gradijentni spust* - (engl. *batch gradient descent*), a čini ga varijanta u kojoj se najprije izračunavaju potrebne promjene težina δw , koje nastaju svim uzorcima u skupu za učenje ili grupi (engl. *batch*), nakon čega se vrši ažuriranje težina. Prema literaturi, takav je pristup često brži u konvergenciji i stabilniji, ali je s druge strane podložniji zapinjanju u lokalnim optimumima te uz njega nije moguće koristiti metode drugog reda prilikom učenja konvolucijskih neuronskih mreža [85]; te

– *stohastički gradijentni spust* - (engl. *stochastic gradient descent*) tj. varijanta gradijentnog spusta kod koje se nakon svakog uzorka ažuriraju težine.

S obzirom na ažuriranje težina nakon prolaza svakog uzorka kroz mrežu, vrijednosti težina više osciliraju te je sama tehnika otpornija na moguća zaglavljivanja u lokalnim optimumima. Nadalje, primjenom ove tehnike za učenje konvolucijskih neuronskih mreža, moguće je dodatno ubrzavanje postupka metodama drugog reda [85].

3.4.2 Algoritam povratnog prostiranja pogreške u konvolucijskim slojevima

Algoritam povratnog prostiranja pogreške je efikasan postupak za ažuriranje svake težine u mreži na način da konačan izlaz bude jednak ciljanom izlazu. U konvencionalnoj višeslojnoj unaprijednoj mreži to je najkorišteniji algoritam za učenje opisan u većem broju radova [86].

Za razliku od unaprijedne umjetne neuronske mreže, konvolucijska neuronska mreža koristeći strategiju dijeljenja težina između neurona izračunava konvolucije nad podacima, pri čemu je sama konvolucijska jezgra formirana pomoću težina. Nakon konvolucije slijedi operacija sažimanja, koja kao oblik nelinearnog uzorkovanja progresivno smanjuje prostornu veličinu reprezentacije ulaznih podataka, čime se smanjuje količina izračunavanja i parametara u mreži.

Između konvolucijskih i slojeva sažimanja nalazi se sloj aktivacijske funkcije, npr. ReLU, koji, kako je navedeno u odjeljku 3.1, ima za cilj uvođenje nelinearnosti u mrežu.

Nakon nekoliko izmjena konvolucijskih i slojeva sažimanja, veličina mape značajki se smanjuje i izdvajaju se kompleksnije značajke. Na kraju, sadržaj mape značajki šalje se u potpuno povezani višeslojni perceptron, kako bi se izvršila klasifikacija. Zadnji sloj višeslojnog perceptrona je izlazni sloj u kojem se definira funkcija gubitka, koja se koristi za određivanje načina i iznosa kojom postupak učenja korigira odstupanje između predviđenih i stvarnih klasa. S obzirom na razlike konvolucijske neuronske mreže u odnosu na klasičnu unaprijednu umjetnu neuronsku mrežu, algoritam povratnog prostiranja pogreške modificiran je za potrebe primjene u konvolucijskim i slojevima sažimanja. Naime, s obzirom na svojstvo dijeljenja težina, konvolucijska jezgra (koja predstavlja težine) utječe na sve neurone u izlaznoj mapi značajki te je prilikom računanja unazadne propagacije potrebno izračunati vrijednosti grešaka, odnosno promjene težina po svim neuronima mape te sumirati njihov utjecaj.

U nastavku, bit će opisan algoritam povratnog prostiranja pogreške u konvolucijskoj neuronskoj mreži opisanoj ulaznim slojem I , jezgrama K i pomacima b .

Izlaz iz konvolucijskog sloja predstavljen je jednadžbom (3.10)

$$(I * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} K_{m,n} \cdot I_{i+m,j+n} + b, \quad (3.10)$$

u kojoj je $I \in \mathbb{R}^{H \times W}$, $K \in \mathbb{R}^{k_1 \times k_2 \times D}$ za skup jezgri D i $b \in \mathbb{R}^D$ za svaku jezgru.

Za sljedeći skup jednadžbi, koristit će se sljedeća notacija:

l - sloj mreže ($l = 1$ je prvi sloj, a $l = L$ je zadnji sloj)

x - ulaz u mrežu dimenzija $H \times W$ sa iteratorima i, j

w - jezgra težina sa dimenzijama $k_1 \times k_2$ sa iteratorima m, n

$w_{m,n}^l$ - matrica težina koje spajaju neurone u sloju $l-1$ sa neuronima u sloju l

b^l - iznos pomaka u sloju l

$x_{i,j}^l$ - konvoluirani ulazni vektor sloja l sa dodanim pomakom

$o_{i,j}^l$ - izlazni vektor sloja l

$f()$ - aktivacijska funkcija

Prema definiciji, nakon konvolucije ulaznih podataka $x_{i,j}^l$ može se definirati kao:

$$x_{i,j}^l = \sum_m \sum_n w_{m,n}^l o_{i+m,j+n}^{l-1} + b^l, \quad (3.11)$$

$$x_{i,j}^l = \text{rot}_{180^\circ} \{ w_{m,n}^l \} * o_{i,j}^{l-1} + b_{i,j}^l x_{i,j}^l, \quad (3.12)$$

a na izlazni signal primijenjena je aktivacijska funkcija kako slijedi iz izraza (3.13). Izraz (3.12) uobičajena je prezentacija procesa konvolucije, pri kojem je isti definiran kao unakrsna korelacija sa dijagonalno zakrenutom jezgrom.

$$o_{i,j}^l = f(x_{i,j}^l) \quad (3.13)$$

U izlaznom sloju mreže izračunava se predviđeni izlaz y_p za ukupni broj klasificiranih uzoraka. S obzirom da je u učenju mreže poznata ciljna vrijednost t_p , moguće je izračunati pogrešku predviđanja odnosno funkciju gubitka. Standardno, koristi se srednje kvadratno odstupanje, prema izrazu (3.14).

$$E = \frac{1}{2} \sum_p (t_p - y_p)^2 \quad (3.14)$$

Učenje se postiže prilagođavanjem težina na način da se smanji funkcija gubitka. U klasičnom pristupu algoritma s povratnim prostiranjem pogreške, težine se prilagođavaju prema smjeru gradijentnog spusta na površini pogreške E .

U povratnom koraku, potrebno je ažurirati težine. S obzirom na prethodno definirano svojstvo dijeljenja težina, tijekom unaprijedne propagacije konvolucijski postupak osigurava da jezgra težina utječe na svaki izlazni podatak. Konvolucija ulazne mape značajki dimenzija $H \times W$ i jezgre težina dimenzija $k_1 \times k_2$ tvori izlaznu mapu značajki dimenzija $(H - k_1 + 1) \times (W - k_2 + 1)$. U prvom povratnom koraku, nastoji se primjenom lančanog pravila izračunati $\frac{\partial E}{\partial w_{m,n}^l}$, odnosno vrijednost utjecaja točno određene težine jezgre $w_{m,n}^l$ na funkciju gubitka E , prema jednadžbi (3.15).

$$\frac{\partial E}{\partial w_{m,n}^l} = \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \frac{\partial E}{\partial x_{i,j}^l} \frac{\partial x_{i,j}^l}{\partial w_{m,n}^l} = \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \delta_{i,j}^l \frac{\partial x_{i,j}^l}{\partial w_{m,n}^l} \quad (3.15)$$

Uvrštavanjem izraza za $x_{i,j}^l$ iz jednadžbe (3.11) i proširivanjem dobiva se izraz (3.16).

$$\frac{\partial x_{i,j}^l}{\partial w_{m,n}^l} = \frac{\partial}{\partial w_{m,n}^l} \left(\sum_m \sum_n w_{m,n}^l o_{i+m,j+n}^{l-1} + b^l \right) \quad (3.16)$$

Daljnja razrada izraza (3.16) i izračun parcijalnih derivacija za sve komponente rezultira nulom za sve komponente osim za $m = m'$ i $n = n'$ u $w_{m,n}^l o_{i+m,j+n}^{l-1}$, kako slijedi u (3.17).

$$\begin{aligned}
\frac{\partial x_{i,j}^l}{\partial w_{m',n'}^l} &= \frac{\partial}{\partial w_{m',n'}^l} \left(w_{0,0}^l o_{i+0,j+0}^{l-1} + \dots + w_{m',n'}^l o_{i+m',j+n'}^{l-1} + \dots + b^l \right) \\
&= \frac{\partial}{\partial w_{m',n'}^l} \left(w_{m',n'}^l o_{i+m',j+n'}^{l-1} \right) \\
&= o_{i+m',j+n'}^{l-1}
\end{aligned} \tag{3.17}$$

Uvrštavanjem izraza (3.17) u (3.15) dobiva se:

$$\begin{aligned}
\frac{\partial E}{\partial w_{m',n'}^l} &= \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \delta_{i,j}^l o_{i+m',j+n'}^{l-1} \\
&= \text{rot}_{180^\circ} \left\{ \delta_{i,j}^l \right\} * o_{m',n'}^{l-1}
\end{aligned} \tag{3.18}$$

Sumiranje po iteratorima i, j rezultat je dijeljenja težina unutar konvolucijskog sloja, a ukupne sume predstavljaju skup svih gradijenata $\delta_{i,j}^l$ svih izlaza sloja l , gdje je:

$$\delta_{i,j}^l = \frac{\partial E}{\partial x_{i,j}^l}, \tag{3.19}$$

a što se može objasniti kao mjera utjecaja promjene podatka $x_{i,j}^l$ na funkciju gubitka E .

$$\begin{aligned}
\frac{\partial E}{\partial x_{i',j'}^l} &= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \frac{\partial E}{\partial x_{i'-m,j'-n}^{l+1}} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l} \\
&= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i'-m,j'-n}^{l+1} \frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l}
\end{aligned} \tag{3.20}$$

Uvrštavanjem lančanog pravila i sumiranjem dobiva se izraz (3.20), a uz dodatno raspisivanje dijela izraza vezanog uz $x_{i'-m,j'-n}^{l+1}$ koji je jednak $\sum_{m'} \sum_{n'} w_{m',n'}^{l+1} o_{i'-m+m',j'-n+n'}^l + b^{l+1}$ dobiva se:

$$\begin{aligned}
\frac{\partial x_{i'-m,j'-n}^{l+1}}{\partial x_{i',j'}^l} &= \frac{\partial}{\partial x_{i',j'}^l} \left(\sum_{m'} \sum_{n'} w_{m',n'}^{l+1} o_{i'-m+m',j'-n+n'}^l + b^{l+1} \right) \\
&= \frac{\partial}{\partial x_{i',j'}^l} \left(\sum_{m'} \sum_{n'} w_{m',n'}^{l+1} f \left(x_{i'-m+m',j'-n+n'}^l \right) + b^{l+1} \right).
\end{aligned} \tag{3.21}$$

Izračunom parcijalnih derivacija za sve komponente dolazi se do vrijednosti nula za sve komponente osim za komponente gdje je $m' = m$ i $n' = n$. Time dio izraza $f \left(x_{i'-m+m',j'-n+n'}^l \right)$ postaje $f \left(x_{i',j'}^l \right)$, a $w_{m',n'}^{l+1}$ postaje $w_{m',n'}^{l+1}$. Uvrštavanjem korigirane jednadžbe (3.21) u jednadžbu (3.20) dolazi se do konačnog izraza:

$$\frac{\partial E}{\partial x_{i',j'}^l} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i-m,j-n}^{l+1} w_{m,n}^{l+1} f'(x_{i',j'}^l). \quad (3.22)$$

S obzirom na prethodno definirano dijagonalno rotiranje jezgre (3.12) u unaprijednom koraku, u unazadnom koraku također se konvolucija prezentira kao unakrsna korelacija sa dijagonalno rotiranom matricom jezgre, kako je prikazano u (3.23).

$$\begin{aligned} \frac{\partial E}{\partial x_{i',j'}^l} &= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i-m,j-n}^{l+1} w_{m,n}^{l+1} f'(x_{i',j'}^l) \\ &= \text{rot}_{180^\circ} \left\{ \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta_{i+m,j+n}^{l+1} w_{m,n}^{l+1} \right\} f'(x_{i',j'}^l) \\ &= \delta_{i',j'}^{l+1} * \text{rot}_{180^\circ} \left\{ w_{m,n}^{l+1} \right\} f'(x_{i',j'}^l) \end{aligned} \quad (3.23)$$

3.4.3 Algoritam povratnog prostiranja pogreške u slojevima sažimanja

Pri korištenju algoritma povratnog prostiranja pogreške u slojevima sažimanja, potrebno je propagirati pogrešku iz mape trenutnog sloja u mapu prethodnog sloja. S obzirom na prethodno definirano svojstvo nepostojanja parametara za učenje kod sažimajućih slojeva, oni pri povratnom prostiranju pogreške isključivo prosljeđuju pogrešku unatrag. Način prosljeđivanja pritom ovisi o definiranom tipu sloja sažimanja.

Ukoliko se koristi sažimanje aritmetičkom sredinom, u unaprijednom koraku svaki neuron mape sudjeluje u definiranju mape idućeg sloja. Iz tog razloga, potrebno je osigurati da se i pogreška prostire unazad do svakog neurona na način da svaki neuron zaprima jednaku vrijednost pogreške definiranu Kroneckerovim produktom jedinične matrice i matrice pogreške trenutnog sloja.

Sažimanje maksimalnom vrijednošću podrazumijeva unaprijednu propagaciju samo maksimalne vrijednosti prethodne mape značajki. Nužno je zapisati vrijednost lokacije korištenih maksimuma, jer u povratnom koraku pogreška propagira do točno određenog neurona prethodnog sloja.

3.5 Normalizacija grupe

Ulazi u neurone jednog sloja određeni su aktivacijskim odzivima neurona prethodnog sloja, a s obzirom da prilikom učenja dolazi do promjene distribucija tih aktivacija, otežano je učenje težina neurona u sljedećim slojevima. Naime, težine neurona u sljedećim slojevima moraju se prilagoditi svojstvima novih distribucija. Postupkom grupe normalizacije postiže se smanjenje

promjenjivosti distribucija aktivacija neurona unutar neuronske mreže. U konvolucijskim neuronskim mrežama to je moguće postići minimiziranjem kovarijantnog pomaka⁴ kroz normalizaciju aktivacije konvolucijskih slojeva za pojedini neuron nad svim elementima pojedine mape značajki, što posljedično omogućava učenje na stabilnoj distribuciji ulaznih podataka.

Sam postupak se sastoji od transformiranja ulaza neurona $x_1, x_2 \dots x_n$ unutar grupe n na sljedeći način:

$$\hat{x}_i = \frac{x_i - \mu_G}{\sqrt{\sigma_G^2 + \varepsilon}}, \quad (3.24)$$

gdje su μ_G i σ_G^2 aritmetička sredina i varijanca grupe, a ε čini mali broj koji osigurava da je nazivnik različit od nule. Normalizacijom prema jednadžbi (3.24) ograničava se ekspresivna moć pripadnog neurona, a da bi se to izbjeglo uvode se, kako je prikazano u jednadžbi (3.25). parametri γ i β sa svojstvom učenja, koji mogu ograničiti ili pojačati normalizirane vrijednosti \hat{x}_i prema potrebama mreže.

$$y_i \leftarrow \gamma \hat{x}_i + \beta \quad (3.25)$$

U razmatranom sloju najprije se, prema jednadžbi (3.24), normalizira aktivacija svakog kanala oduzimanjem aritmetičke sredine mini-grupa μ_G i dijeljenjem sa varijancom mini-grupe σ_G^2 .

Potom unutar sloja algoritam korigira (pomiče) ulaz, prema jednadžbi (3.25), tj. pomoću ispravljivog odmaka β i faktora razmjera γ omogućuje učenje u slučaju da aritmetička sredina i varijanca sa iznosom nula nisu optimalni za sljedeći sloj. β i γ su parametri za učenje koji se ažuriraju tijekom učenja mreže.

Kada završi učenje mreže, u sloju normalizacije grupe izračunava se srednja vrijednost i varijanca preko cijelog skupa vježbanja i pohranjuje se, kako bi se prilikom korištenja mreže koristili već naučeni parametri za normalizaciju aktivacija.

⁴ Kovarijantni pomak (engl. *covariate shift*) – pojava statistički značajne razlike u distribucijama unutar skupova podataka za vrijeme procesa strojnog učenja

3.6 Regularizacija

Općeniti cilj algoritama umjetne inteligencije svodi se na uspješnu predikciju na novim, dotad za algoritam nepoznatim podacima, a njihova sposobnost za takve operacije naziva se generalizacija. Prilikom učenja, posebice na manjim skupovima podataka, a uslijed velikog kapaciteta učenja može se dogoditi prenaučenosť konvolucijske neuronske mreže, čime se gube svojstva prepoznavanja značajki na novim skupovima podataka odnosno poopćavanja. Naime, pronalazak prave vrijednosti kapaciteta neuronske mreže je čest problem pri dizajniranju hiperparametara. U slučaju u kojem je potrebno povećati svojstva generalizacije umjetne neuronske mreže, može se koristiti pristup u kojem se u funkciju gubitka dodaje regularizacijski član s ciljem penalizacije povećanja složenosti strukture mreže. Praktično, takav regularizacijski član nastoji potisnuti ostale vrijednosti težina prema nuli, što dovodi do isključivanja neurona odnosno smanjenja složenosti modela.

Najčešće se koriste dvije vrste regularizacija, L_1 (3.26) i/ili L_2 (3.27):

$$L_1 = \lambda \|w\|_1 = \lambda \sum_i |w_i| \quad (3.26)$$

$$L_2 = \lambda \|w\|_2^2 = \lambda \sum_i w_i^2 \quad (3.27)$$

gdje w označava vektor težina, w_i i -tu težinu, a λ regularizacijski hiperparametar.

Osim navedenih, kod dubokih neuronskih mreža koristi se i regularizacija temeljena na metodi nasumičnog gašenja neurona prilikom učenja (engl. *dropout*), čime se smanjuje kapacitet mreže za učenje i otežava pretreniranost. Nasumičnost gašenja se pritom definira za pojedine slojeve, uz određenu vjerojatnost za aktiviranje gašenja za pojedini neuron.

3.7 Određivanje hiperparametara modela

Svaki model strojnog učenja sadrži određene parametre, koji se mogu mijenjati. Ti parametri uvelike utječu na ishod postupka učenja, detaljno objašnjenog u odjeljku 3.4. U kontekstu CNN-a, takvim parametrima smatraju se npr. parametar koji određuje optimizator učenja, brzina učenja, veličina i broj jezgri ili parametri koji određuju faktore regularizacije. Ti parametri različiti su od parametara ili težina koje model uči i podešava tijekom treninga, i kao takvi se nazivaju *hiperparametrima*.

Postoje 2 vrste parametara u neuronskim mrežama:

- parametri koje mreža uči (engl. *learnable parameters*) (Ovaj tip parametara mreža uči i podešava za vrijeme treninga pomoću algoritma povratnog prostiranja pogreške, a u njih spadaju težine i pomaci. Broj parametara učenja se povećava s povećanjem broja neurona i slojeva.) i-
- hiperparametri (engl. *hyperparameters*) kao tip parametara, koji se zadaju prije samog procesa učenja, te se ne uče naknadno.

Hiperparametri u konvolucijskim neuronskim mrežama mogu se podijeliti u 3 kategorije [87]:

1. hiperparametri s kontinuiranim realnim brojevima koji mogu poprimiti bilo koju vrijednost unutar određenog raspona, a primjeri takvih hiperparametara su: koeficijent učenja η , momentum, parametar regularizacije λ ;
2. hiperparametri s diskretnim cjelobrojnim vrijednostima: broj konvolucijskih slojeva, broj jezgri u svakom sloju, veličina jezgri, broj epoha i dr.; te
3. hiperparametri koji su diskretne i/ili binarne vrijednosti i sadrže konačan broj mogućnosti: odabir optimizatora, odabir aktivacijske funkcije, metoda smanjivanja koeficijenta učenja (engl. *learning rate decay method*).

S obzirom na to da umjetne konvolucijske neuronske mreže sadrže veći broj hiperparametara čija promjena utječe na ishod učenja, potrebno je prilikom postupka učenja pronaći skup hiperparametara koji daje najbolji ishod, tj. najbolje performanse mreže na testnim skupovima podataka.

Uzevši u obzir prethodno spomenutu činjenicu o većem broju hiperparametara i dodajući pritom vrijeme potrebno za učenje sa jednim odabranim skupom hiperparametara, može se zaključiti kako postupak njihovog prilagođavanja s ciljem osiguravanja performansi na testnom skupu nije trivijalan zadatak.

Na primjer, koeficijent učenja pomaže odrediti u kolikoj mjeri se težine prilagođavaju svakoj iteraciji. Preniska vrijednosti koeficijenta učenja uzrokovat će predugo trajanje učenja da bi se postigli dobri rezultati. S druge strane, previsoki koeficijent učenja uzrokovat će da učenje nikada ne dostigne dobre rezultate jer se težine prilagođavaju željenoj vrijednosti u svakoj iteraciji.

Hiperparametri konvolucijske neuronske mreže navedeni su u nastavku:

- *dubina mreže i broj konvolucijskih slojeva;*
- *broj jezgri* Svaki konvolucijski sloj sadrži otprije odabrani broj jezgri. Broj jezgri utječe na broj mapa značajki u svakom sloju. Težine se u njima mijenjaju kako mreža trenira.;

- *veličina jezgre* Dimenzije jezgre se definiraju širinom i visinom matrice. Jezgre su kvadratne matrice uobičajeno dimenzija 3x3 ili 5x5. Svako polje unutar jezgre predstavlja jednu težinu te se odabirom veličine jezgre odabire i broj težina u njoj. Npr. jezgra dimenzija 5x5 imala bi 25 težina. U slučaju da se izabere premala veličina jezgre, mreža možda neće biti sposobna prenijeti sve bitne informacije u procesu konvolucije, dok prevelika jezgra usporava proces učenja.;
- *dopunjavanje* (engl. *padding*) Povećavaju se dimenzije ulazne matrice dodajući nule oko rubova kako bi se dobila mapa značajki jednakih dimenzija kao i ulazna matrica. Kod ovog se hiperparametra odlučuje hoće li se dopunjavanje koristiti ili neće.;
- *korak s* (engl. *stride*) Broj za koji se je potrebno pomaknuti u vertikalnom i horizontalnom smjeru kod operacija konvolucije i sažimanja. U slučaju korištenja 1D signala, pomicanje se vrši samo u horizontalnom smjeru.;
- *redoslijed i broj slojeva* Empirijski je pokazano da mreže s više slojeva zahtijevaju mnogo manji broj neurona da bi postigle iste rezultate i obično generaliziraju bolje na temelju nepoznatih podataka [87].;
- *koeficijent učenja η* Najčešće se odabire u iznosu između 0,1 do 0,0001, a smatra se jednim od ključnih hiperparametara, jer o njemu ovisi konvergencija mreže.;
- *aktivacijska funkcija* Neke od najpoznatijih su: sigmoidalna, tangens hiperbolna aktivacijska funkcija, zglobna aktivacijska funkcija, itd.;
- *broj epoha* Epohu sačinjava jedan prolaz svih uzoraka kroz neuronsku mrežu za vrijeme treninga. Nakon svake epohe se može vidjeti pogreška koju mreža daje. Manjim brojem epoha se skraćuje vrijeme treniranja, ali i pogoršava istreniranost mreže.;
- *veličina mini grupe podataka* (engl. *batch size*) O ovom hiperparametru ovisi koliko će uzoraka u jednom prolazu proći kroz mrežu. Dobar izbor veličine mini grupe znatno ubrzava učenje mreže, ne gubeći previše na njenoj točnosti. Ukoliko se koristi algoritam gradijentni spust s mini grupama, težine će se ažurirati nakon svake mini grupe.
- *parametri regularizacije* Odabir iznosa λ ukoliko se radi o regularizaciji L1 ili L2 vektorskim normama. Osim L1 i L2 regularizacije, može se koristiti i tehnika isključivanja (engl. *dropout*) neurona tijekom učenja, no generalno se kod konvolucijskih neuronskih mreža češće koristi tehnika normalizacije grupe (engl. *batch normalization*), koja se pokazuje efektivnijom.;

- *optimizacijska tehnika* Osim standardnog odabira gradijentnog spusta sa momentumom, moguće je kao optimizator učenja koristiti i alternative kao ADAM (engl. *adaptive moment estimation*) ili RMSProp (engl. *root mean square propagation*), i dr.
- *funkcija gubitka* Ovisno o problemu (klasifikacija ili regresija) i ulaznim podacima odabire se funkcija gubitka. Unakrsna provjera se najčešće koristi za klasifikaciju, a za regresiju se koriste: srednja kvadratna pogreška, korijen srednje vrijednosti kvadrata pogreške, srednja apsolutna pogreška, i dr., te
- *sloj sažimanja* Odabire se kojom će tehnikom algoritam obavljati sažimanje nad mapama značajki i veličina matrice $n_K * n_K$ koja će vršiti sažimanje. Najčešće korištene tehnike su: sažimanje maksimumom i sažimanje uprosječavanjem.

Postoje optimalni hiperparametri za svaki zadatak, tako da je za osiguravanje najboljih rezultata modela važno pronaći takav optimalni skup hiperparametara.

Optimizacijom hiperparametara nastoji se odabrati skup hiperparametara koji će najbolje utjecati na učenje neke neuronske mreže. Problem optimizacije hiperparametara se pomoću optimizacijskih metoda nastoji premjestiti s čovjeka na računalo jer je čovjek sklon pogreškama i nailazi na teškoće pri shvaćanju višedimenzionalnih problema. Optimizacija hiperparametara spada u skup problema crne kutije (engl. *black-box problem*) jer se povezanost izbora hiperparametara s izvedbom mreže koja koristi te hiperparametre ne može opisati matematičkom formulom. Pravilan izbor hiperparametara ima ključnu ulogu u smanjenju vremena učenja i poboljšanju točnosti mreže.

Hiperparametre može mijenjati i analitičar, ali vrlo je mala vjerojatnost da će postići optimalan izbor hiperparametara, pogotovo ako se radi o višedimenzionalnom problemu. Trenutno, u području održavanja postoje radovi u kojima se kao postupak određivanja parametara navodi ručno podešavanje hiperparametara. Takvo podešavanje hiperparametara je ostvarivo, ali je vrlo neučinkovito i uobičajeno se temelji na dobrim početnim parametrima iz prethodne literature. Stoga se za određivanje hiperparametara danas uobičajeno koriste računalne tehnike.

3.8 Računalne tehnike optimizacije hiperparametara modela

Postoje četiri najpoznatije računalne metode za optimiziranje hiperparametara neuronskih mreža:

- mrežno pretraživanje (engl. *grid search*),
- nasumično pretraživanje (engl. *random search*),

- genetski algoritam (engl. *genetic algorithm*),
- Bayes optimizacija (engl. *Bayesian optimization*).

3.8.1 Mrežno pretraživanje

Mrežno pretraživanje je najjednostavnija metoda optimizacije hiperparametara. Svodi se na pretraživanje funkcije u određenom intervalu te praćenje i odabir maksimalne vrijednosti funkcije $f(x)$ za svaki x . Svaka vrijednost x je jednako udaljena od prethodne i sljedeće. Što se uzme manji interval između x -ova, to će mreža imati veću vjerojatnost doći blizu maksimuma, ali će za posljedicu imati dulje trajanje učenja. U slučaju optimizacije hiperparametara mreže svaki x bi predstavljao moguću vrijednost za samo jedan hiperparametar. U slučaju odabira više od jednog hiperparametara, to bi značilo da se ne traži samo najviša vrijednost x -a nego najveća vrijednost funkcije svih ostalih hiperparametara. Valja imati na umu da utjecaj hiperparametara na mrežu zavisi o vrijednostima svih ostalih hiperparametara. Ako se odabire između n hiperparametara, tada to postaje problem mrežnog pretraživanja jer on postaje n -dimenzionalan. Kako bi se dobio dovoljno dobar rezultat potrebno je izvesti povećani broj ispitivanja. Mrežno pretraživanje garantira optimalan pronalazak zadanih hiperparametara, no s direktnom posljedicom velikog utroška vremenskog resursa jer je potrebno izračunati sve kombinacije.

3.8.2 Nasumično pretraživanje

Metoda nasumičnog pretraživanja je slična metodi mrežnog pretraživanja, ali može u kraćem vremenu dati bolje rezultate jer se vrijednosti točaka odabiru nasumično, umjesto da se odabiru u pravilnim razmacima, što bitno smanjuje vrijeme trajanja procesa. Rizik ove metode je u tome što algoritam, za razliku od algoritma mrežnog pretraživanja, ne posjećuje sve točke pa postoji rizik da nasumične točke nisu ni blizu maksimuma. U multidimenzijskim problemima, metoda nasumičnog pretraživanja daje bolje rezultate uz kraći vremenski ciklus od prethodno definirane metode mrežnog pretraživanja.

3.8.3 Genetski algoritam

Genetski algoritam je metaheuristička metoda inspirirana prirodom i prirodnom selekcijom. Pripada skupini evolucijskih algoritama. Ukratko, genetski algoritam oponaša prirodu tako što omogućuje opstanak najboljih primjeraka.

Da bi došlo do evolucije u prirodi moraju biti ispunjena tri zahtjeva:

- nasljednost – roditelji prenose svoje gene na djecu;

- varijacija – organizmi unutar populacije posjeduju različita svojstva i
- konkurencija – rađa se više djece nego što ih preživi.

Prevedeno u jezik optimizacije hiperparametara neuronske mreže, organizmi bi bili modeli, a njihova svojstva (geni) različite kombinacije hiperparametara. Roditelj će biti model. Kad se model kombinira s drugim modelom nastaje novi model koji dijeli svojstva ta dva modela – njihovo dijete.

Algoritam optimizacije hiperparametara temeljen na genetskom algoritmu djeluje na sljedećim principima:

1. *inicijalizacija* Kreira se skup rješenja zvan populacija. Svako rješenje unutar populacije posjeduju različita svojstva. U optimizaciji hiperparametara ta svojstva predstavljaju hiperparametre.;
2. *vrednovanje rješenja* Kod neuronskih mreža modeli se evaluiraju pomoću: točnosti (engl. *accuracy*), preciznosti (engl. *precision*), odziva (engl. *recall*), F1-mjere (engl. *F1-score*);
3. *odabir najboljih rješenja na temelju mjera za evaluaciju*, tj. modeli s najboljom kombinacijom hiperparametara;
4. *najbolja rješenja se križaju međusobno kako bi se dobila slična rješenja*. Svako rješenje dobiveno iz križanja prijašnjih rješenja nosi barem neka svojstva od svojih *roditelja*;
5. *nekim od novih rješenja se nasumično mijenjaju određena svojstva*, tj. vrijednosti određenih hiperparametara Ovaj se proces naziva mutacijom. Mutacije se provode kako bi se smanjila vjerojatnost konvergencije modela u lokalni minimum.; te se
6. *nova rješenja uspoređuju sa starim rješenjima*, tj. svojim *roditeljima* te se brišu ona koja su lošija. Svaka iteracija predstavlja jednu generaciju.

Ovaj proces se ponavlja dok se ne dođe do zadovoljavajućeg rješenja.

Ovaj način optimizacije je računalno zahtjevan na velikim skupovima podataka, tj. zahtjeva veći broj iteracija kako bi ostvario potencijal konvergencije prema optimalnom rješenju. Nadalje, sam algoritam temeljen na evolucijskom pristupu sadrži veći broj hiperparametara (funkcija gubitka, veličina populacije itd.), a odabir kojih također znatno utječe na ishod i trajanje optimizacije.

3.8.4 Bayesova optimizacija

Kao što je već spomenuto, problem optimizacije hiperparametara je u tome što je vrednovanje funkcije cilja, kako bi se pronašao optimalan skup, izuzetno računalno intenzivno. Svaki put

kada se testira različite hiperparametre, mora se učiti model na trening podacima, testirati model na testnom skupu, a zatim izračunati metriku evaluacije. S velikim brojem hiperparametara i složenim modelima kao što su ansambli modela ili duboke neuronske mreže koje mogu trenirati danima, ovaj proces brzo postaje jako teško ručno izvediv. Bayesova optimizacija je razumniji način pronalaženja optimalne kombinacije hiperparametara u odnosu na mrežno i nasumično pretraživanje jer uzima u obzir prijašnje rezultate na temelju kojih odabire iduću kombinaciju hiperparametara koja će biti ispitana. Naime, mrežno i nasumično pretraživanje uopće ne obraćaju pažnju na prošle rezultate i nastavljaju pretraživati cijeli raspon hiperparametara, iako je možda jasno da je optimalna vrijednost u malom području. Sam proces odabira sličan je ljudskom donošenju odluka, pri čemu se analiziraju prethodna iskustva na temelju kojih se donosi odluku. Ova metoda je osobito korisna u slučajevima kad je izračun funkcije cilja dugotrajan ili računalno intenzivan. Bayesova optimizacija u konačnici skraćuje vrijeme trajanja procesa fokusiranjem na nešto duljem i smislenijem načinu izbora kombinacije hiperparametara koje će mreža učiti, što dovodi do broja iteracija učenja.

Bayesova optimizacija može se definirati kao proces izrade vjerojatnosnog modela funkcije cilja te njegove upotrebe za odabir najboljih hiperparametara, koji će biti evaluirani na stvarnoj funkciji cilja kroz proces testiranja naučene mreže. Dakle, sam proces može se u osnovi podijeliti na fazu istraživanja te fazu pomaka u smjeru najvećeg dobitka.

$$x_{opt} = \arg \min_{x \in D} f(x) \quad (3.28)$$

Matematički, to se može prikazati jednadžbom (3.28), gdje $f(x)$ predstavlja rezultat evaluacije mreže naučene na skupu hiperparametara na testnom skupu podataka, x_{opt} je skup hiperparametara koji se koriste, a x može poprimiti bilo koju vrijednost u domeni D .

Bayesov pristup, za razliku od mrežnog ili nasumičnog pretraživanja, pohranjuje i koristi rezultate prošlih evaluacija za formiranje vjerojatnosnog modela, pri čemu se korišteni hiperparametri mapiraju sa vjerojatnošću rezultata na funkciji cilja, tj. $P(\text{rezultat evaluacije} | \text{hiperparametri})$.

Ovaj model u literaturi je često nazivan i surogat model funkcije cilja i prezentiran sa $p(y|x)$. Surogat je mnogo lakše optimizirati od stvarne funkcije cilja, a Bayesove metode rade tako da za izračun stvarne funkcije cilja koriste hiperparametre koji najbolje funkcioniraju na surogatnoj funkciji.

Gaussovim procesom oblikuje se surogat model koji za cilj ima aproksimiranje funkcije cilja. Surogat model procjenjuje funkciju cilja na nizu točaka u domeni. Nakon svake iteracije ta se procjena ažurira. Nakon što se funkcija cilja izračuna u nekoj točki x , model se ažurira uslijed novih informacija koje je izvukao iz te točke. Što je više točaka surogat modela izračunano, surogat model bolje aproksimira funkciju cilja.

Osnovne značajke potrebne za provedbu Bayesove optimizacije su integracija modela Gaussovog slučajnog procesa (engl. *Gaussian process*, GP) kao *priornog* modela i izračun funkcije stjecanja (engl. *acquisition function*). Prema tome, može se zaključiti kako se Bayesov optimizator sastoji od dvije glavne komponente:

- statističkog modela za izračun kvalitete korištene vrijednosti hiperparametra, i
- stjecajne funkcije za izračun točaka sljedeće iteracije.

Statistički model za oblikovanje funkcije cilja je Gaussov slučajni proces, a takav model služi za izračun posteriorne Bayesove distribucije vjerojatnosti, koja opisuje potencijalne iznose $f(x)$ u nekoj točki x , tj. radi procjenu $f(x)$ prikazujući kolika je nesigurnost da je ta procjena ispravna. U konačnici, kao što je prikazano jednadžbom (3.28), nastoji se minimizirati funkcija cilja $f(x)$, pri čemu je x unutar ograničene domene.

Pojednostavljenu rutinu algoritma Bayesove optimizacije moguće je opisati pseudokodom (vidi tablicu 3.1).

Tablica 3.1 Standardni pseudokod Bayesopt algoritma

1:	funkcija BAYESOPT(funkcija f , model g , stjecajna funkcija a , inicijalni skup zapažanja I_0)
2:	Postavi <i>prior</i> funkciju Gaussovog slučajnog procesa na f
3:	Izračunaj f za n_0 točaka. Postavi $n = n_0$
4:	Za $n = 1, 2, \dots, n_0$
5:	Izračunaj $x_n = \arg \max_x a(x; g, I_{n-1})$
6:	Evaluiraj $f(x_n)$ za generiranje zapažanja y_n
7:	Ažuriraj $I_n = I_{n-1} \cup (x_n, y_n)$
8:	Ažuriraj model g koristeći I_n
9:	završi petlju

10: Vрати rješenje: točka sa najvećim $f(x)$ ili točka sa najvećom vrijednosti aritmetičke sredine posteriorne funkcije

11: **završi funkciju**

Prior modelom Gaussovog slučajnog procesa sa dodanim Gausovim šumom opisuje se pretpostavljeni način izgleda stvarne funkcije. Gaussov proces je stohastički proces (skup slučajnih varijabli indeksiranih vremenom ili prostorom), takav da svaki konačni skup tih slučajnih varijabli ima multivarijantnu normalnu distribuciju, tj. svaka konačna linearna kombinacija varijabli normalno se distribuira. Kao i ostale tehnike nadziranog strojnog učenja, Gaussov proces zahtijeva označeni skup podataka za učenje u izradi modela. Ako je $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ skup uzoraka za učenje iz nekog procesa, pri čemu je $y = \{y_i, i = 1, \dots, N\}$ skup ciljnih vrijednosti, a $X = \{x_i, i = 1, \dots, N\}$ je $d \times n$ matrica odgovarajućih koordinatnih točaka ulaznog uzorka u \mathbb{R}^d , svaka vrijednost ciljne varijable y_i može biti predstavljena kroz transformacijsku funkciju $f(x_i)$, kao što je prikazano u jednadžbi:

$$y_i = f(x_i) + \varepsilon_i, \quad (3.29)$$

gdje je ε_i Gaussov šum sa poznatom varijancom σ_n^2 .

Funkcija distribucije priora Gaussovog slučajnog procesa $f(x)$ je određena vektorom srednjih vrijednosti $\mu(x; \theta)$ i kovarijacijskom matricom $k(x_i, x_j | \theta)$, gdje je θ vektor parametara kernela [88]. Model Gaussovog procesa reprezentira ciljnu funkciju kao neparametarsku distribuciju specificiranu skupom trening uzoraka D i kovarijantnom funkcijom. Naime, u nadziranom učenju, očekuje se da će točke sa sličnim ulaznim vrijednostima imati slične vrijednosti ciljne varijable. Stoga, ključna ideja korištenja Gaussovog procesa jest da su vrijednosti funkcija u različitim točkama korelirane, a vrijednost kovarijance između dviju vrijednosti funkcija u različitim točkama $f(x_i)$ i $f(x_j)$ ovisi o ulaznim podacima x_i i x_j . Ta ovisnost može biti prikazana kovarijantnom funkcijom ili kernelom $k(x_i, x_j)$. Kovarijantna funkcija može biti opisana različitim funkcijama kernela i parametrizirana uz pomoć vektora parametara θ te u tom slučaju vrijedi $k(x_i, x_j | \theta)$. Jedna od često korištenih kernel funkcija je funkcija automatskog određivanja relevantnosti (engl. *automatic relevance determination*, ARD), prema jednadžbi

$$k(x_i, x_j | \theta) = \sigma_f^2 \exp\left[-\frac{1}{2} \frac{(x_i - x_j)^T (x_i - x_j)}{l^2}\right], \quad (3.30)$$

gdje je l karakteristična skala duljine (engl. *characteristic length-scale*), koja opisuje ugladenost funkcije. Skale duljine određuju koliko daleko mogu biti ulazne vrijednosti x kako bi vrijednosti izlaza postale nekorelirane. Za potrebe ovog istraživanja, koristit će se ARD Matérn 5/2 kernel funkcija, prema prijedlogu iz [89] i jednadžbe:

$$k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right), \quad (3.31)$$

gdje je r Euklidska udaljenost između točaka x_i i x_j te se može izračunati preko jednadžbe (3.32):

$$r = \sqrt{(x_i - x_j)^T (x_i - x_j)} \quad (3.32)$$

Porastom broja točaka, kako se prikupljaju podaci, ova se distribucija ažurira tako da se formira *posterior* nad mogućim funkcijama koje su konzistentne s promatranim podacima.

Na taj način, moguće je Bayesovim zaključivanjem, kroz kontinuirano ažuriranje surogatnog vjerojatnosnog modela, postići efektivno minimiziranje pogreške na stvarnoj funkciji cilja.

Stjecajna funkcija je pomoćna funkcija s ciljem procjene kvalitete sljedeće točke x iz surogat modela, a koja će se ispitati u funkciji cilja. Takvom funkcijom se definira ravnoteža između istraživanja novih područja u prostoru funkcije cilja i iskorištavanja vrijednosti iz područja koja su već ispitana. Dakle, stjecajna funkcija evaluira kvalitetu točke x na temelju posteriorne funkcije distribucije Q (tj. trenutno najboljeg rješenja). Točka sa najvećom vrijednošću funkcije stjecanja uzima se kao točka na kojoj će se učiti u sljedećoj iteraciji. Intuitivno, može se pretpostaviti kako je potrebno maksimizirati $\sigma_Q(x)$ ili $-\mu_Q(x)$. Maksimiziranje $\sigma_Q(x)$ potaknulo bi pretraživanje dalje od prethodnih promatranih mjesta. Problem s ovim pristupom je u tome što će se pretraživanja nastaviti i nakon što je model siguran u lokaciju globalnog minimuma. Što se tiče maksimiziranja $-\mu_Q(x)$, takva procedura predstavlja ozbiljan rizik od zapadanja u prostor rješenja koji možda nije ni lokalni optimum [90] budući da nema poticaja za istraživanje na područjima koja su daleko od trenutnog najboljeg rješenja. Postoji veći broj funkcija stjecanja, a najčešće se odabiru:

- a. vjerojatnost poboljšanja (engl. *probability of improvement*)

Stjecajna funkcija vjerojatnosti poboljšanja izračunava vrijednosti x_{best} , odnosno lokacije sa najnižom vrijednošću aritmetičke sredine posteriora te $\mu_Q(x_{best})$ tj. najnižu vrijednost aritmetičke sredine posteriora. Nakon toga, izračunava se vjerojatnost da će nova točka x voditi ka boljoj vrijednosti funkcije cilja za vrijednost parametra m .

$$PI(x, Q) = P_Q(f(x) < \mu_Q(x_{best}) - m) \quad (3.33)$$

- b. donji stupanj povjerenja (engl. *lower confidence bound*, UCB)

Korištenje navedene stjecajne funkcije podrazumijeva izračun funkcije $G(x)$, prema jednadžbi:

$$G(x) = \mu_Q(x) - 2\sigma_Q(x), \quad (3.34)$$

gdje je $G(x)$ krivulja koja spušta granicu aritmetičke sredine posteriora za iznos $2\sigma_Q$ u svakoj točki. Bayesov optimizator potom maksimira negativnu vrijednost G , kao što je prikazano u jednadžbi (3.35).

$$LCB = 2\sigma_Q(x) - \mu_Q(x) \quad (3.35)$$

- c. očekivano poboljšanje (engl. *expected improvement*)

Stjecajne funkcije u ovoj kategoriji procjenjuju očekivani iznos poboljšanja funkcije cilja, ignorirajući pritom vrijednosti koje uzrokuju samo povećanje cilja. Dakle, ukoliko je x_{best} lokacija sa najnižom vrijednošću aritmetičke sredine posteriora, a $\mu_Q(x_{best})$ najniža vrijednost aritmetičke sredine posteriora, očekivano poboljšanje može se izraziti kao u jednadžbi (3.36).

$$EI(x, Q) = E_Q \left[\max(0, \mu_Q(x_{best}) - f(x)) \right] \quad (3.36)$$

U ovom radu, za potrebe odabira hiperparametara modela konvolucijske neuronske mreže za ranu procjenu uzroka kvara opreme implementiran je Bayesov algoritam optimizacije sa stjecajnom funkcijom očekivanog poboljšanja, koja je dokazana kao efikasna [89] u brzom dohvaćanju globalnog optimuma.

3.9 Sažetak poglavlja

Na temelju istraživanja literature definirane su i detaljno opisane konvolucijske neuronske mreže, njihova arhitektura i slojevi te je u koracima objašnjen princip učenja istih. Određen je termin hiperparametara te su dane osnovne značajke najvažnijih hiperparametara korištenih za

oblikovanje mreže u tijeku učenja. Nadalje, odjeljak 3.4 opisuje proces učenja konvolucijske neuronske mreže. Dodatno, na temelju izvedenih zaključaka o važnosti odabira ispravnih vrijednosti hiperparametara, navedene su i ukratko definirane osnovne tehnike optimizacije hiperparametara. Objasnjena je ispravnost korištenja Bayesove optimizacije za potrebe optimizacije hiperparametara konvolucijskih neuronskih mreža te je navedena tehnika i detaljno opisana.

4. EKSPERIMENTALNI POSTAV I MJERNA OPREMA

Eksperimentalni postav korišten za pokusni dio ovog rada, koji je proveden u *Laboratoriju za održavanje Fakulteta strojarstva i brodogradnje*, prikazan je na slici 4.1.



Slika 4.1 Eksperimentalni postav

Eksperimentalni postav sastoji se od:

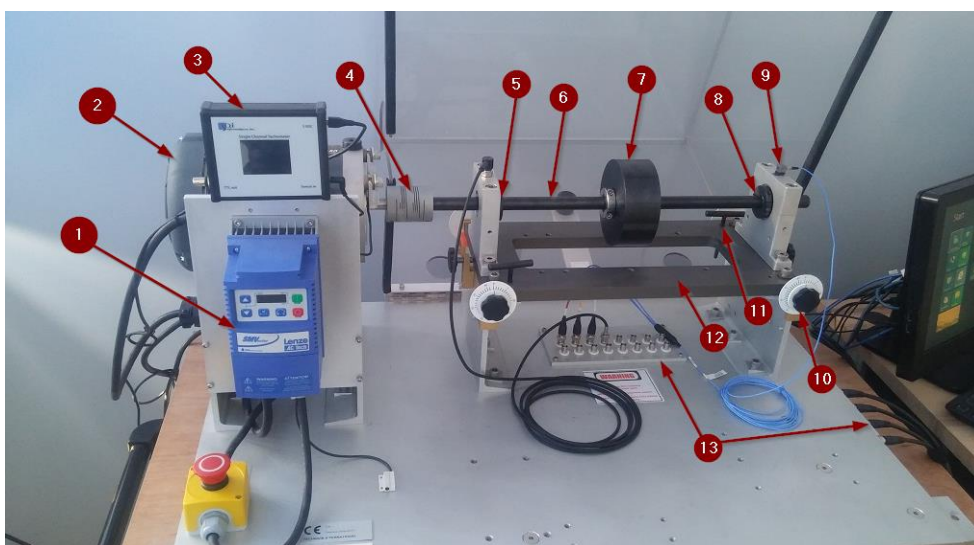
- simulatora kvarova rotacijske opreme proizvođača *SpectraQuest* sa modulima za simuliranje različitih tipova kvarova,
- beskontaktnog optičkog senzora učestalosti vrtnje,
- troosnog senzora akceleracije proizvođača *PCB Piezotronics*,
- sustava za prikupljanje podataka *National Instruments cDAQ-9191* sa modulom za akviziciju podataka sa senzora vibracija,
- mjernog računala za obradu i pohranu prikupljenih podataka, i
- računala za razvoj modela.

4.1 Simulator kvarova rotacijske opreme

Simulator kvarova rotacijske opreme tipa *Expert* proizvođača *SpectraQuest* dio je laboratorijskog postava *Laboratorija za održavanje Fakulteta strojarstva i brodogradnje*, a radi se o uređaju sklopljenom u sklopu inovacijskog projekta iz programa *Provjere inovativnog koncepta* (engl. *proof of concept*) *Inteligentni računalni sustav rane dijagnostike kvarova*

rotacijskih elemenata strojeva, u cijelosti ostvarenog unutar *Laboratorija za održavanje Fakulteta strojarstva i brodogradnje*. Uređaj takvog tipa korišten je u istraživanjima [59,91–95] za razumijevanje različitih vibracijskih obrazaca kroz provođenje kontroliranih eksperimenata, pri čemu je simuliran rad rotacijske opreme u ispravnom stanju te u različitim stanjima kvarova, a podaci prikupljeni sa montiranih senzora služili su za daljnju analizu.

Simulator kvarova (slika 4.2) pokreće asinhroni motor snage 0,75 kW preko *Lenze SMVector* frekvencijskog pretvarača opremljenog regulatorom brzine vrtnje. Frekventni regulator omogućuje regulaciju učestalosti vrtnje u rasponu od 0 – 6 000 o/min promjenom izlazne frekvencije napajanja elektromotora, a trenutna učestalost vrtnje mjeri se integriranim optičkim senzorom koji generira diferencijalne TTL signale na izlaznom sučelju. Snaga se preko elastične spojke prenosi na vratilo uležišteno sa dva kuglična ležaja tipa *ER12K* montirana koncentričnim sigurnosnim prstenom. Postolje simulatora te osnovni moduli pričvršćeni su vijcima, na način kojim se omogućuje postavljanje dodatnih modula odnosno, simuliranje različitih vrsta kvarova koji pritom generiraju različite obrasce vibracijskih opterećenja. Osnovne tehničke karakteristike simulatora kvarova dane su u tablici 4.1.



1. Frekventni regulator *Lenze SMVector*;
2. Asinhroni motor,
3. Ekran s prikazom učestalosti vrtnje,
4. Spojka,
5. Prednji ležaj,
6. Vratilo,
7. Uteg,
8. Stražnji ležaj,
9. troosni IEPE akcelerometar,
10. Vijak za horizontalno poravnavanje osi,
11. Vijak za vertikalno poravnavanje osi,
12. Postolje,
13. BNC priključnice

Slika 4.2 Simulator kvarova

Tablica 4.1 Osnovne tehničke karakteristike simulatora kvarova

Elektromotor	trofazni, 0,75 kW, napajanje jednom fazom napona 220 V
Regulacija	Frekventni regulator <i>Lenze SMVector</i>
Učestalost vrtnje rotora	Varijabilna u rasponu 0 – 6 000 o/min
Vratilo	d=19,05 mm (3/4 ")
Ležajevi	Dva kotrljajuća ležaja tipa <i>ER12K</i> u aluminijskom horizontalno odvojivom kućištu za olakšanu montažu/demontažu. Moguća montaža u 5 različitih raspona.
Priključci za mjernu opremu	16 BNC priključnica unutar zaštitnog plašta, povezanih sa vanjskim panelom za povezivanje akvizicijskog sustava
Postolje	Portabilno aluminijsko postolje s mogućnošću horizontalnog i vertikalnog poravnavanja osi.
Masa	150 kg

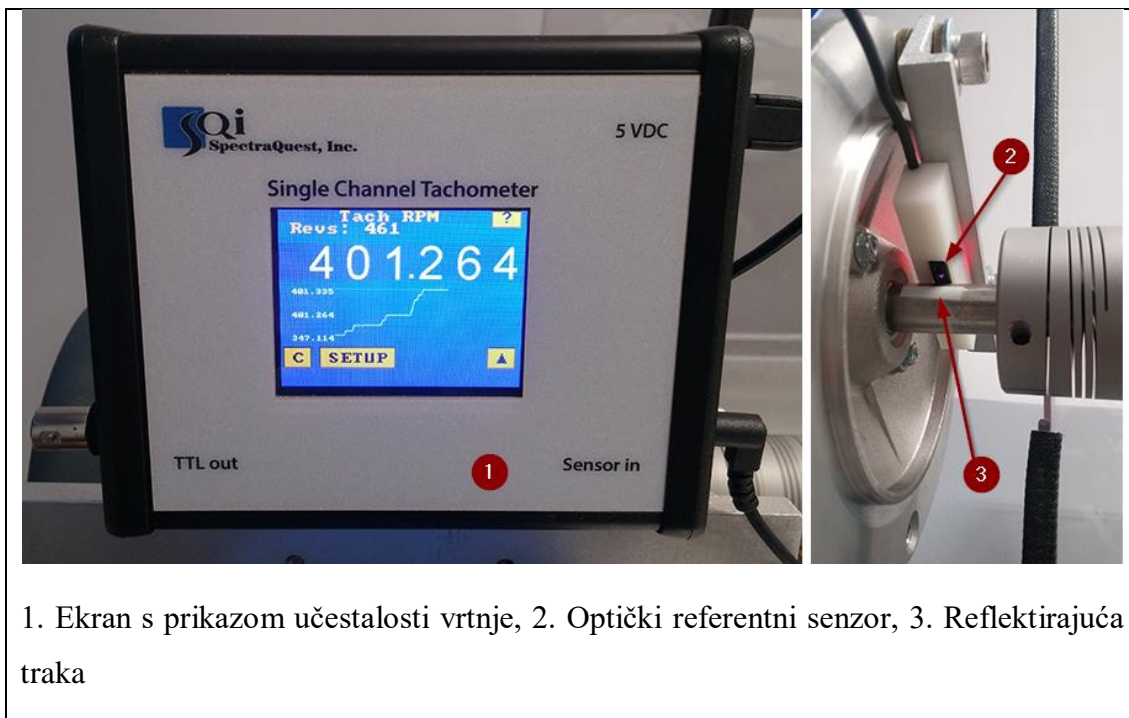
Osnovna konstrukcija simulatora kvarova podržava jednostavnu modifikaciju i promjene s ciljem primjene u različitim vrstama istraživanja.

4.2 Mjerni sustav eksperimentalnog postava

Prikupljanje procesnih podataka pri izvođenju eksperimenta izvodi se uz pomoć mjerne opreme, koja je za potrebe eksperimenta montirana na simulator kvarova.

4.2.1 Beskontaktni optički senzor učestalosti vrtnje

Na vratilo je sa strane elektromotora zalijepljena reflektirajuća traka, a na postolje je montiran optički senzor. Optički senzor radi na principu generiranja digitalnog impulsa za svako očitavanje reflektirajuće trake (1 impuls/okretaj). Iako danas postoje i sofisticiraniji načini mjerenja učestalosti vrtnje (npr. kroz upotrebu enkodera), s obzirom na prirodu eksperimenta i korištenje konstantne učestalosti vrtnje pri samom izvođenju, u ovom slučaju koristi se jednostavna, a dovoljno točna metoda zasnovana na optičkom principu. Na slici 4.3 vidljiv je instalirani senzor spojen na ekran sa prikazom učestalosti vrtnje.



Slika 4.3 Ekran i optički senzor

4.2.2 Senzor akceleracije

Vibracijski signali sa simulatora prikupljeni su mjernim sustavom sastavljenim od piezoelektričnog tro-osnog senzora akceleracije s integriranim pojačalom (engl. *integrated electronic piezoelectric*, IEPE) proizvođača *PCB Piezotronics* tipa *356B21* (vidi sliku 4.4) spojenog na tri ulazna kanala na sustav za prikupljanje i obradu podataka. Piezoelektrični tip senzora na izlazu generira naboj proporcionalan deformaciji kristala. Taj se naboj pretvara u napon nabojskim pojačalom koje je ovdje integrirano u sam senzor. Time se sprečava postupno zamicanje radne točke (*drift*) koji je inače prisutan i teško u cijelosti otklonjiv kod svih piezoelektričnih senzora koji nisu IEPE tipa. Senzor se napaja strujom iz akvizicijskog sučelja prilagođenom za IEPE senzore na način da se oklop koaksijalnog kabla ostavlja na 0 V, dok se kroz jezgru koaksijalnog kabla preko veznog kondenzatora dovodi napajanje. Vezni kondenzator se stavlja između analognog ulaza i točke gdje se dovodi napajanje te na taj način sprečava da analogni ulaz mjeri napon napajanja senzora.



Slika 4.4 Senzor akceleracije PCB 356B21

Tehničke karakteristike tro-osnog akcelerometra dane su u tablici proizvođača (vidi tablicu 4.2)

Tablica 4.2 Osnovne tehničke karakteristike senzora akceleracije

	Iznos	Mjerna jedinica
Mjerno područje akceleracije	± 500	g
Osjetljivost	10	mV/g
Frekvencijsko područje	2 – 7 000	Hz
Rezonantna frekvencija	≥ 55	kHz
Pojasna rezolucija (1 do 10 000 Hz)	0,04	m/s ² rms
Nelinearnost	≤ 1	%
Napon pobude	18 - 30	VDC
Konstantna struja pobude	2 - 20	mA
Izlazni napon	7 - 12	VDC

4.2.3 Sustav za prikupljanje i obradu podataka

Za potrebe eksperimenta korišten je sustav za prikupljanje i obradu podataka temeljen na *National Instruments CompactDAQ* platformi. Takav sustav sastoji se od osnovne jedinice (postolja) spojenog sa mjernim računalom te modula za spajanje senzora. Za potrebe razmatranog eksperimenta, odabran je modul *NI 9234* za dinamičku akviziciju analognog signala, specifično namijenjen IEPE sensorima. Modul se sastoji od 4 ulazna kanala sa BNC

priključnicama. Svaki kanal opremljen je podesivim izvorom konstantne struje koji se može prilagoditi karakteristikama priključenog IEPE senzora. *Modul NI 9234*, čije su tehničke karakteristike dane u tablici 4.3, se koristi u eksperimentu zajedno sa *CompactDAQ NI 9191* šasijom, a koja služi kao sklop za okidanje i prijenos podataka između *NI 9234* modula i mjernog računala.

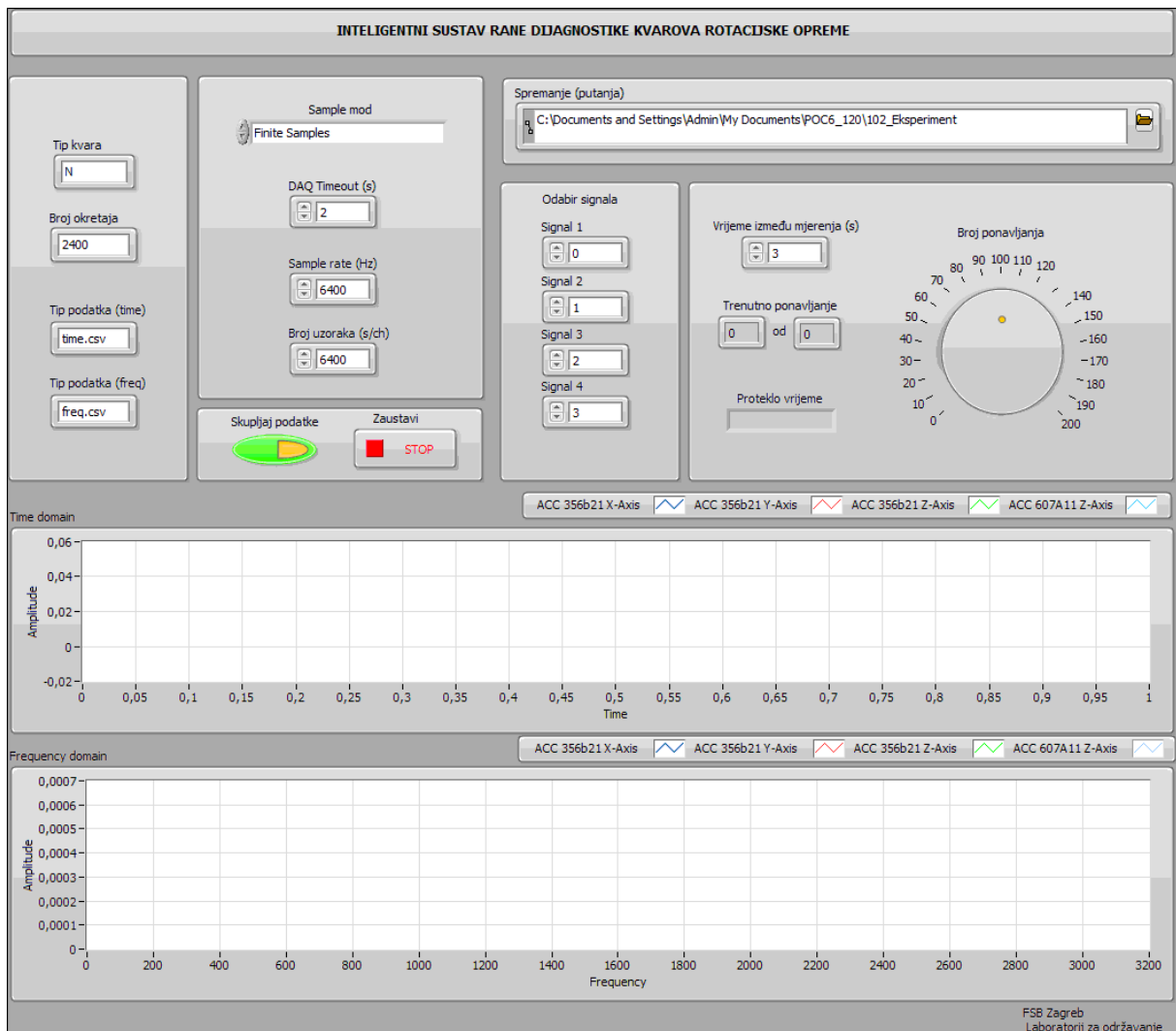
Tablica 4.3 Tehničke karakteristike NI 9234 modula

Raspon signala	Broj kanala	Brzina uzorkovanja	Povezivost	ADC rezolucija	Tip ADC
± 5 V	4	51,2 kS/s/ch	4 x BNC	24 bita	Delta-Sigma

4.2.4 Korisničko sučelje mjernog računala

Podešavanje parametara sustava prikupljanja i obrade podataka te nadzor prikupljanja i obrade podataka regulirani su uz pomoć mjernog računala. Za potrebe prikupljanja i obrade podataka, razvijena je mjerna aplikacija u sklopu *LabView* okruženja. Uz pomoć grafičkog korisničkog sučelja razvijene aplikacije (slika 4.5), odabire se:

- način prikupljanja podataka,
- frekvencija uzorkovanja,
- broj uzoraka u svakom mjerenju,
- broj ponavljanja uzorkovanja,
- vrijeme između mjerenja, i
- parametri potrebni za generiranje i pohranu mjernih datoteka.



Slika 4.5 Razvijeno grafičko sučelje aplikacije za nadzor tijekom eksperimenta

Dodatno, korisničkim sučeljem navedene aplikacije za nadzor tijekom eksperimenta, ručnim unosom zadaju se parametri stanja pokusa, kao što su:

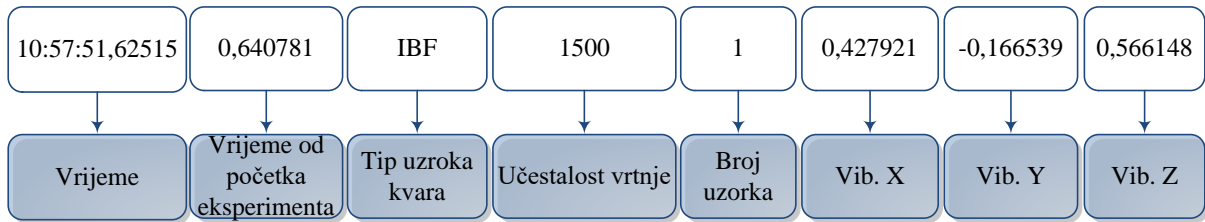
- tip simuliranog uzroka,
- učestalost vrtnje rotora, te
- trenutni broj iteracije uzorkovanja,

koji se prilikom generiranja mjerne datoteke zapisuju uz izmjerene vrijednosti akceleracije u X, Y i Z smjeru.

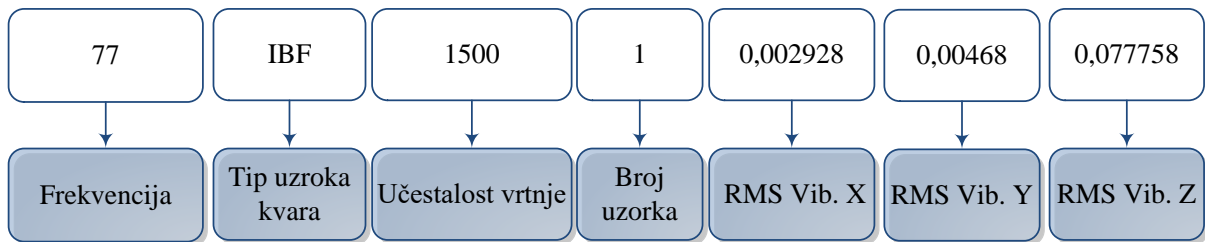
Na temelju unesenih vrijednosti, pokretanjem aplikacije generiraju se i otvaraju dvije tekstualne datoteke za pohranu izmjerenih vrijednosti, s točnom određenim nazivljem koje odgovara parametrima stanja pokusa (za detaljniji opis vidi slike 4.6 i 4.7):

- a) tekstualna datoteka za pohranu vremenske serije sa sufiksom *time* i ekstenzijom *csv* i .

b) tekstualna datoteka za pohranu podataka iz frekvencijske domene sa sufiksom *freq* i ekstenzijom *csv*.



Slika 4.6 Primjer strukture podataka u vremenskoj domeni



Slika 4.7 Primjer podataka u frekvencijskoj domeni

Datoteke ostaju otvorene za vrijeme izvođenja eksperimenta, a po završetku mjerenja *.csv* datoteka se zatvara i pohranjuje na tvrdi disk.

5. PLANIRANJE POKUSA I MJERENJE

Navedeno poglavlje obuhvaća detaljnu razradu plana pokusa, snimanje, postupke obrade signala te način pohrane datoteka u međuspremnik za razmjenu sa sučeljem za inteligentnu obradu.

Planiranje glavnog eksperimenta izvedeno je u fazama. Nakon obavljenog pregleda literature i izvedenih preliminarnih pokusnih mjerenja, definirani su tipovi uzroka kvarova, koji će se simulirati. Nastavno, na osnovu preporuke proizvođača te izvedenih prepokusa, utvrđeni su parametri simulatora kao i način odvijanja samog eksperimenta.

Snimanje signala vibracija tijekom izvođenja glavnog eksperimenta izvedeno je korištenjem senzora akceleracije i sustava za prikupljanje podataka te mjernog računala za pohranu opisanih u odjeljku 4.2. Broj mjerenja i parametri eksperimenta vezani uz frekvenciju prikupljanja te broj uzoraka u svakom mjerenju određeni su na temelju analize dostupnih preporuka proizvođača, mogućnostima laboratorijskog mjernog sustava te niza probnih mjerenja izvedenih sa različitim parametrima tijekom prepokusa.

5.1 Odabir modula simuliranih tipova uzroka kvarova

Simulator kvarova rotacijske opreme platforma je za eksperimentiranje i simuliranje vibracijskih obrazaca različitih stanja odnosno kvarova opreme. Međutim, za potrebe detaljnijih istraživanja i izučavanje uzročno-posljedičnih veza pojedinih tipova kvarova, potrebno je koristiti dodatne module, čije djelovanje u konačnici može opisati ponašanje opreme pri zadanom tipu kvara.

Za potrebe provođenja eksperimenta, korišteni su nadogradni moduli za simulaciju pojedinih tipova kvarova:

- rotori za simulaciju neravnoteže,
- ekscentričan rotor,
- nagnuti rotor i
- oštećen kotrljajući ležaj:
 - ležaj s oštećenjem unutarnje staze ležaja,
 - ležaj s oštećenjem vanjske staze ležaja,
 - ležaj s oštećenjem kuglice, te
 - ležaj s kombiniranim oštećenjima.

5.1.1 Rotori za simuliranje neravnoteže

Vibracije, a potom i oštećenja i kvarovi nastali uslijed debalansiranog rotacijskog tijela spadaju u kategoriju najčešćih neispravnih stanja rotacijske opreme. Stanje neravnoteže ili debalansa može se definirati kao ono stanje u kojem rotacijska oprema emitira vibracijsko opterećenje nastalo kao rezultat centrifugalne sile, tj. stanje nejednolike distribucije mase oko osi rotacije tijela. Pritom se mogu definirati dva terminološka pojma: glavna os inercije i geometrijska os rotacije. Glavna os inercije predstavlja onu os, oko koje bi rotor rotirao da nije ograničen ležajevima (engl. *principle inertia axis*, PIA), dok je geometrijska os rotacije stvarna, fizikalna os rotacije rotora (engl. *geometric centerline*, GCL). U slučaju podudarnosti dvaju osi, ne postoji fizikalni temelj za postojanje vibracija uslijed djelovanja centrifugalne sile. U drugom slučaju, mogu se definirati tri scenarija neravnoteže opreme uslijed nepodudarnosti spomenutih osi:

- statička neravnoteža – glavna os rotacije i geometrijska os rotacije su paralelne,
- neravnoteža uslijed momenta – glavna os rotacije i geometrijska os rotacije se sijeku u centru horizontalne udaljenosti ležišta, te
- dinamička neravnoteža – glavna os rotacije i geometrijska os rotacije su nekoincidentne i neparalelne (postoji istovremena statička neravnoteža i neravnoteža momenta).

Na simulator kvarova mogu se nadograditi dva rotora jednakih promjera sa dva reda provrta sa navojem, u koje je moguće pri simuliranju odgovarajućeg tipa neravnoteže vijcima učvrstiti dodatnu masu, koja će pri vrtnji uzrokovati vibracijski obrazac. Rotori se pričvršćuju na vratilo spojnicom sa vijkom. Izgled rotora za simuliranje debalansa prikazan je na slici 5.1.



Slika 5.1 Rotor za simulaciju debalansa

5.1.2 Ekscentričan rotor

Uz neravnotežu, ekscentričnost rotacijskog elementa opreme također je čest uzrok vibracija kvarova. Ekscentricitetom rotirajućeg tijela definira se stanje odmaknutosti linije centra rotacije zupčanika, prijenosnika, armature motora ili bilo kojeg rotora od geometrijske linije rotacije vratila.

Za simulaciju ekscentričnog rotacijskog tijela, simulator kvarova proširiv je sa nadogradnim modulom u obliku diska s asimetrično pozicioniranim centrom rotacije (slika 5.2). Prilikom rotacije takvog tijela, dolazi do generiranja specifičnog vibracijskog obrasca koje uzrokuju sile uslijed asimetrične rotacije tijela.



Slika 5.2 Modul ekscentričnog rotora

5.1.3 Nagnuti rotor

Nagnutost rotirajućeg tijela također se smatra karakterističnim uzrokom vibracija i rezultira pojavom kvarova. Nagnutim rotirajućim tijelom definira se stanje odmaknutosti linije centra rotacije zupčanika, prijenosnika, armature motora ili bilo kojeg rotora od geometrijske linije rotacije vratila.

Za simuliranje nagnutog rotacijskog tijela, simulator kvarova proširiv je sa nadogradnim modulom u obliku diska s provrtom (slika 5.3), čija je os vrtnje zamaknuta $0,5^\circ$ od osi rotacije vratila. Prilikom rotacije takvog tijela, dolazi do generiranja specifičnog vibracijskog obrasca koje uzrokuju sile uslijed rotacije nagnutog rotirajućeg tijela.



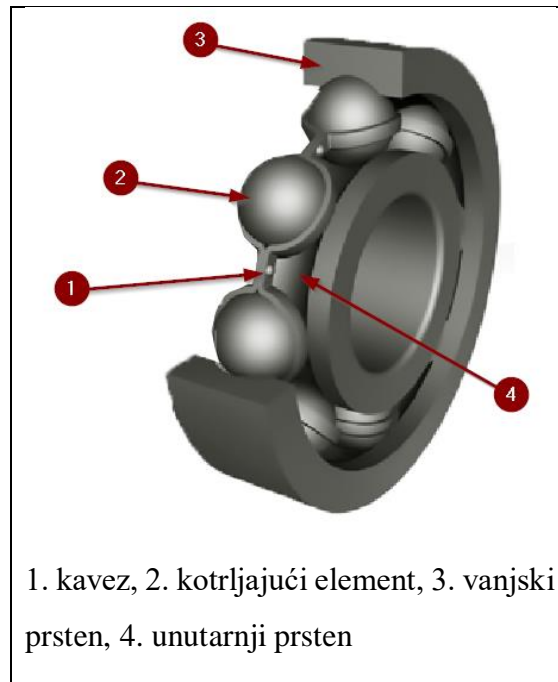
Slika 5.3 Modul nagnutog rotora

5.1.4 Oštećeni kotrljajući ležajevi

Rotacijska oprema koristi ležajeve za prijenos opterećenja i osiguravanje odnosa između stacionarnih i rotirajućih elemenata, pri čemu se većinom koriste kotrljajući ležajevi. Iako je za taj element strojeva moguće matematički procijeniti vijek trajanja, kotrljajući ležajevi skloni su preuranjenim kvarovima pa tako jedva 10 % od njih dosegne izračunati vijek trajanja L10, odnosno očekivani životni vijek ostalih 90 % sličnih ležajeva izvrnutih sličnim radnim uvjetima [96]. Pritom se kao uzroci preuranjenih kvarova mogu smatrati pogreške u podmazivanju i montaži, a također se mogu izdvojiti i umor materijala, korozija i brineling (trajno utiskivanje radne površine ležaja uslijed djelovanja sile). Svaki od navedenih tipova uzroka kvarova ležaja stvara primarno oštećenje ležaja, a koje posljedično dovodi do sekundarnih oštećenja koja uzrokuju kvar.

Zbog njihove nužnosti i vitalnog doprinosa operativnoj funkcionalnosti većine rotirajuće opreme, nužna je pravovremena odluka o zamjeni ležaja.

Kotrljajući ležaj sastoji se od četiri osnovne komponente: unutarnjeg prstena, vanjskog prstena, kotrljajućeg elementa i kućišta (kaveza) (slika 5.4), pri čemu primarno oštećenje nastalo nekim od prije spomenutih uzroka može nastati na svakom od njih.



Slika 5.4 Glavni elementi kotrljajućeg ležaja

S obzirom na navedeno, postoje 4 različite dominantne frekvencije, prema kojima je analizom moguće utvrditi postojanje oštećenja na pojedinom elementu:

- frekvencija unutarnje staze kotrljanja (engl. *ball pass frequency inner race*),
- frekvencija vanjske staze kotrljanja (engl. *ball pass frequency outer race*),
- frekvencija kotrljajućeg elementa (engl. *ball pass frequency*), te
- frekvencija kaveza (engl. *fundamental train frequency*).

U ovom radu, korišteni su, uz ispravne, 4 ležaja tipa *ER12K* sa primarnim oštećenjima na pojedinim elementima, kako slijedi (slika 5.5):

- ležaj sa oštećenjem unutarnje staze kotrljanja unutarnjeg prstena,
- ležaj sa oštećenjem vanjske staze kotrljanja vanjskog prstena,
- ležaj sa oštećenjem kotrljajućeg elementa (kuglice), te
- ležaj sa kombiniranim oštećenjem pojedinih elemenata.



1. Kombinirano oštećenje, 2. oštećenje kotrljajućeg elementa, 3. oštećenje unutarnjeg prstena, 4. oštećenje vanjskog prstena

Slika 5.5 Ležajevi s primarnim oštećenjima

5.1.5 Ostale korištene nadogradnje

Uz u prethodnim odjeljcima navedene nadogradne module simulatora kvarova, dodatno je korišten inercijski disk mase 5 kg s ciljem povećanja osnovnog opterećenja na ležajeve, odnosno povećanja spektralne amplitude cijelog sustava. Inercijski disk (slika 5.6) se pričvršćuje na utvrđenu poziciju na vratilu dvjema spojnica.



Slika 5.6 Inercijski disk za vratilo

Na temelju definiranih korištenih nadogradnih modula i definiranih tipova uzroka kvarova rotacijske opreme, utvrđen je ulazni skup tipova uzroka kvarova za provedbu eksperimenta:

- I. neravnoteža rotora (IMRF),

- II. ekscentričnost rotora (ERF),
- III. nagnutnost rotora (CRF),
- IV. oštećenje vanjske staze kotrljanja kotrljajućeg ležaja (ORBF),
- V. oštećenje unutarnje staze kotrljanja kotrljajućeg ležaja (IRBF),
- VI. oštećenje kotrljajućeg elementa kotrljajućeg ležaja (BBF) te
- VII. kombinirano oštećenje kotrljajućeg ležaja (CBF)

5.2 Izbor parametara eksperimenta

Pod pojmom parametri eksperimenta, podrazumijeva se učestalost vrtnje rotora simulatora, frekvencija uzorkovanja, vrijeme uzorkovanja, vrijeme između mjerenja te broj ponovljenih mjerenja tijekom eksperimenta.

Za potrebe izvođenja glavnog eksperimenta, izveden je veći broj predpokusa s ciljem definiranja korektnih parametara eksperimenta. Proizvođač simulatora navodi moguće korištenje simulatora u području učestalosti vrtnje 0 – 6 000 o/min. S obzirom na korištenje nadogradnih modula, tehničke karakteristike sustava za prikupljanje podataka i troosnog senzora akceleracije te planirani veći broj ponovljenih uzorkovanja, određene su eksperimentalne učestalosti vrtnje od 1 000 o/ min i 1 500 o/min.

Parametri frekvencije uzorkovanja, vremena uzorkovanja i vremena između mjerenja su odabrani s obzirom na tehničku specifikaciju troosnog akcelerometra (vidi tablicu 4.2) te sustava za prikupljanje podataka (vidi tablicu 4.3).

Konačno, nakon niza probnih mjerenja, na temelju dobivenih rezultata odabrana su stanja pokusa, prikazana u tablici 5.1. Postavljeno je ukupno šesnaest kombinacija za snimanje vibracijskog signala unutar glavnog eksperimenta, točnije dvije razine učestalosti vrtnje za normalno (ispravno) stanje opreme te za svaki definirani tip uzroka kvara (IMRF, ERF, CRF, IRBF, ORBF, BBF, CBF). Svaka kombinacija mjerenja ponavljana je 1 500 puta, što je konačnici rezultiralo sa 24 000 ponovljenih mjerenja (uzoraka).

Tablica 5.1 Parametri stanja eksperimenta

Mjerenje	Učestalost vrtnje	Broj ponavljanja mjerenja (broj uzoraka)	Stanje opreme
	n min^{-1}	N /	/
1		1 500	Normalno stanje (NS)
2		1 500	Neravnoteža rotora (IMBF)
3		1 500	Ekscentričnost rotora (ERF)
4		1 500	Nagnutost rotora (CRF)
5	1 000	1 500	Oštećenje vanjske staze kotrljanja kotrljajućeg ležaja (ORBF)
6		1 500	Oštećenje unutarnje staze kotrljanja kotrljajućeg ležaja (IRBF)
7		1 500	Oštećenje kotrljajućeg elementa kotrljajućeg ležaja (BBF)
8		1 500	Kombinirano oštećenje kotrljajućeg ležaja (CBF)
9		1 500	Normalno stanje (NS)
10		1 500	Neravnoteža rotora (IMBF)
12		1 500	Ekscentričnost rotora (ERF)
12		1 500	Nagnutost rotora (CRF)
13	1 500	1 500	Oštećenje vanjske staze kotrljanja kotrljajućeg ležaja (ORBF)
14		1 500	Oštećenje unutarnje staze kotrljanja kotrljajućeg ležaja (IRBF)
15		1 500	Oštećenje kotrljajućeg elementa kotrljajućeg ležaja (BBF)
16		1 500	Kombinirano oštećenje kotrljajućeg ležaja (CBF)

5.3 Snimanje i filtriranje signala

Kvalitetno snimanje signala procesa prvi je i nužan uvjet za uspješno provođenje klasifikacije uzroka kvara rotacijske opreme uz pomoć vibracijskih signala senzora akceleracije. U okviru trenutnog istraživanja, u tu svrhu korišten je mjerni sustav detaljnije opisan u odjeljku 4.2. Konverzija analognog vibracijskog signala u digitalnu reprezentaciju odvija se uz prikupljanje

određenog broja uzoraka prikupljenih u točkama vremena. Premali broj uzoraka ima za posljedicu netočnu digitalnu reprezentaciju stvarnog signala. Posljedično, bilo je potrebno osigurati frekvenciju uzorkovanja višu od maksimalne frekvencije, koja je iskorištena u daljnjoj obradi.

5.3.1 Snimanje ispitnih uzoraka signala

Mjereni signali vibracija u X, Y i Z osi snimani su senzorom vibracija u kraćim vremenskim intervalima, a snimljena mjerenja potom su pohranjena za fazu automatskog izdvajanja značajki.

Jednim mjerenjem se za svaki ispitni uzorak prikuplja 25 600 podataka frekvencijom uzorkovanja $f_s = 51\,200$ Hz. U predpokusima, korištene su različite vrijednosti frekvencije uzorkovanja, a odabrana frekvencija uzorkovanja predstavlja maksimalno moguću za prethodno opisani akvizicijski modul.

5.3.2 Filtriranje signala niskopropusnim filterom

Filtriranje signala s ciljem eliminacije fenomena *alias*-inga, odnosno pridavanje visokofrekvencijskih spektralnih amplituda signalima nižih frekvencija uobičajen je korak prije daljnje obrade vibracijskog signala. Naime, prema Nyquistovom teoremu [97], za točnu interpretaciju točnog sinusnog vala, signal ne smije sadržavati frekvencije koje su više od polovice frekvencije uzorkovanja.

$$f_{Nyquist} = \frac{f_s}{2} \quad (5.1)$$

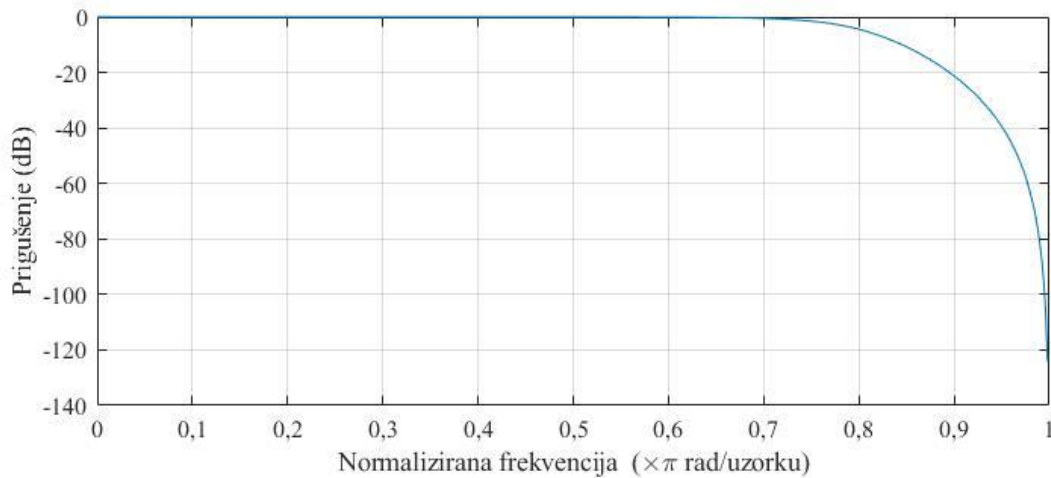
Pri odabiru značajki filtera, u praktičnoj upotrebi često se koristi izraz

$$f_s = f_{max} \cdot 2,56 \quad (5.2)$$

pri čemu se superponira Nyquistovo načelo. S obzirom da idealni frekvencijsko-selektivni filteri nisu praktično provedivi, uslijed značajki kao što su tranzicijski pojas i roll-off efekt, potrebno je koristiti frekvenciju uzorkovanja veću od teorijski zadane Nyquistove frekvencije uzorkovanja.

U procesu filtriranja signala niskopropusnim filtrom primijenjen je niskopropusni Butterworth filter šestog reda sa beskonačnim impulsnim odzivom (engl. *infinite impulse response*, IIR), s ciljem propuštanja signala na frekvenciji nižoj od zadane granične frekvencije. Opća funkcija

filtera prikazana je slikom 5.7. Zadana granična frekvencija iznosi $f_{\max} = 7\ 000\ \text{Hz}$, dok je frekvencija uzorkovanja $f_s = 51\ 200\ \text{Hz}$.



Slika 5.7 Opća funkcija Butterworth filtera trećeg reda

Zadana granična frekvencija ujedno je i gornja granica osjetljivosti senzora vibracija. Snimljeni i filtrirani signali pohranjeni su u vremenskoj domeni, ali i frekvencijskom području. Za pretvorbu u frekvencijsku domenu korišten je standardni FFT algoritam, a pri pohrani podataka sve vrijednosti iznad granične frekvencije troosnog akcelerometra ($f_{\max} = 7\ 000\ \text{Hz}$) izuzete su iz mjerne datoteke.

Za potrebe istraživanja, uveden je termin *vibracijske slike*, kojom je definiran skup podataka sa senzora vibracija obuhvaćen jednim mjerenjem.

Za svako mjerenje odnosno svaki ispitni uzorak dobivene su *vibracijske slike* različitih dimenzija u dvije domene:

- *vibracijske slike* vremenske domene dimenzija $25\ 600 \times 1 \times 3$, u kojima 25 600 predstavlja broj prikupljenih signala u jednom mjerenju, 1 je broj korištenih akcelerometara, a 3 je broj osi akcelerometra; i
- *vibracijske slike* frekvencijske domene dimenzija $1\ 750 \times 1 \times 3$, u kojima 1 750 predstavlja broj pojaseva širine 4 Hz ravnomjerno raspoređenih u mjernom području senzora od 0 - 7 000 Hz za koje su računane prosječne vrijednosti amplituda, 1 je broj korištenih akcelerometara, a 3 je broj osi akcelerometra. Vibracijska slika frekvencijske domene podrazumijeva pretprocesiranje signala.

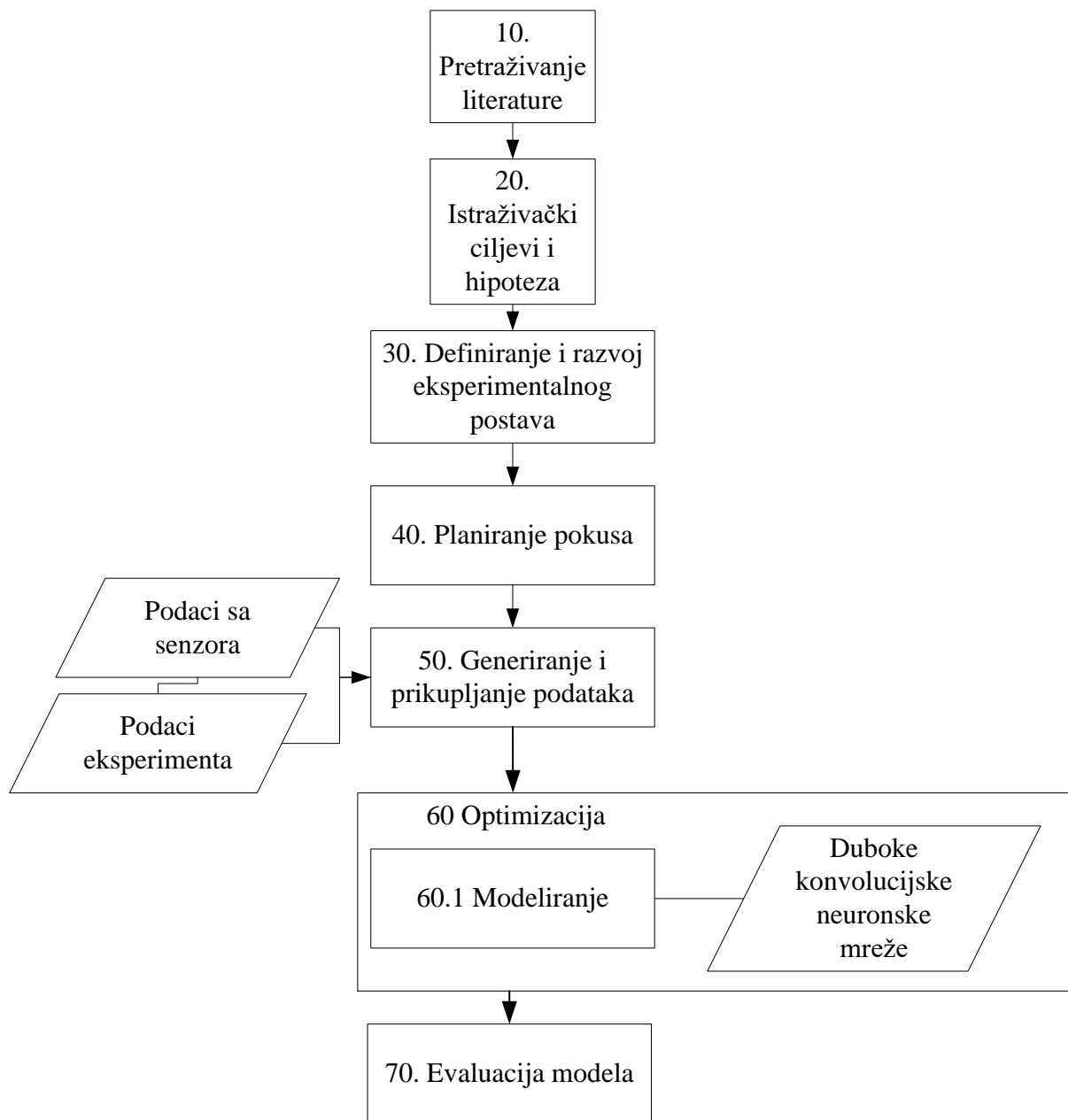
6. MODEL RANE PROCJENE UZROKA KVAROVA ROTACIJSKE OPREME

Rotacijska oprema, kao jedna od najčešće korištenih klasa industrijske opreme, sastoji se od tri glavna dijela: rotora, kotrljajućih ili kliznih ležaja i postolja. Često je smatrana kritičnom opremom proizvodnih sustava, a svoju primjenu pronalazi u velikom broju različitih industrija. Rotacijska oprema može poprimiti značajnu složenost s obzirom na tip prigona, konstrukciju i izvedbu, a nije ju neuobičajeno koristiti u otežanim eksploatacijskim uvjetima. Primjena u takvim uvjetima posljedično znači i veću osjetljivost na različite tipove kvarova. Degradacija svojstava rotacijske opreme može uzrokovati kvar opreme s ozbiljnim sigurnosnim, ekonomskim, ali i ekološkim posljedicama. Moderni pristupi održavanju uokvireni u strategiju prediktivnog održavanja po stanju temelje aktivnosti održavanja na procjeni stanja i otkrivanju potencijalnih uzroka kvarova u ranoj fazi. Svrha održavanja rotacijske opreme najčešće se temelji na sprečavanju kvarova opreme i smanjenju troškova održavanja smanjenjem nepotrebnih sati održavanja.

Vodeći se time, važno je ispravno odabrati način na koji će se provesti ispitivanje trenutnog stanja opreme, odnosno koristiti tehniku koja će imati sposobnost u ranoj fazi prepoznati pojavu potencijalnog uzroka kvara.

Blok shema istraživanja prikazana je na slici 6.1.

Vibrodijagnostika je široko korištena tehnika održavanja po stanju u održavanju rotacijske opreme, zbog, kako je objašnjeno u odjeljku 2.3.1, generalno dobrih performansi pri ranom otkrivanju uzroka kvarova. Ipak, u većini slučajeva teško je dati apsolutne tolerancije vibracija. Stoga postoji očigledna opasnost lošeg prosuđivanja te ljudsko iskustvo i tumačenje izmjerenih vrijednosti i dalje imaju važnu ulogu u praćenju stanja vibracija. S druge strane, s obzirom da vibrodijagnostika podrazumijeva ekspertnu analizu prikupljenih mjerenja, takav proces isto tako unosi subjektivnost pri procjeni stanja rotacijske opreme.



Slika 6.1 Proces istraživanja

Nadalje, s porastom kompleksnosti rotacijske opreme, raste i broj mogućih kvarova, a podsustavi takve opreme kao što su ležajevi ili zupčanci postaju teško dostupni za vizualni ili taktilni pregled zbog samih dimenzija opreme, dugog vremena potrebnog za pripremu pregleda ili okolišnih ograničenja [98].

Iz ovih razloga, nužno je za kritičnu rotacijsku opremu razvijati modele, koji će osigurati veću točnost pri procjeni stanja i ranoj identifikaciji kvara i na taj način prevenirati pojavu nepredviđenih kvarova.

Uzrok promjene razine vibracija rotacijske opreme najčešće je povezan sa promjenom stanja, tj. degradacijom pojedinih dijelova opreme. Trenutni modeli usredotočeni su na otkrivanje odnosno identifikaciju uzroka promjene razine vibracija, odnosno na temelju otkrivanja implicitnog znanja u podacima metodama strojnog učenja. Mogućnost procjene uzroka promjene razine vibracija znači i mogućnost djelovanja odnosno planiranja aktivnosti održavanja.

U odjeljku 2.4 opisan je trenutni stadij razvoja prediktivnog održavanja po stanju rotacijske opreme, a u odjeljku 2.5 obrazložene su umjetne neuronske mreže kao najčešće primjenjivana tehnika umjetne inteligencije u području inteligentnog prediktivnog održavanja po stanju. Dodatno je istaknuto kako se većina konvencionalnih AI modela kao što su BPNN, RBFNN i SOM može primijeniti u svrhu otkrivanja trenutnih informacija o pogreškama i rješavanja problema dijagnostike, kada se radi o manjoj količini podataka i kad modeliranju prethodi proces izdvajanja značajki. Međutim, prilikom rada s velikom količinom sirovih podataka, konvencionalni modeli nemaju adekvatnih mogućnosti procjene stanja i dijagnostike kvarova.

Prema najnovijim istraživanjima, moguće je razviti modele temeljene na konceptima dubokog učenja, kako bi se omogućila kvalitetnija procjena stanja i dijagnostika kvarova rotacijske opreme i za slučajeve kad se radi sa većom količinom industrijskih podataka. U odjeljku 2.7 opisana su istraživanja sa trenutnim nastojanjima u pronalasku adekvatnih modela za procjenu stanja i dijagnostike kvarova u kojima su korišteni koncepti dubokog učenja. Svaka od predloženih metoda pridonijela je i postigla ciljeve u relevantnim eksperimentima. Ipak, svaka od predloženih metoda radi s ograničenim skupom dostupnih podataka, pa u većini slučajeva točnost procjene identifikacije uzroka kvara u opremi još uvijek nije dovoljna za implementacijske svrhe.

Na temelju trenutnih spoznaja, definirani su ciljevi rada i hipoteza.

U ovom poglavlju, bit će predloženi modeli za ranu identifikaciju potencijalnih uzroka kvarova temeljeni na obradi velike količine sirovih podataka. Na temelju dosadašnjih istraživanja i spoznaja (detaljnije opisanih u poglavlju 2), te opisanog načina rada konvolucijske neuronske mreže (poglavljje 3), razvijen je eksperimentalni postav predstavljen u poglavlju 4 te su prikupljeni i pohranjeni podaci prema planu pokusa opisanom u poglavlju 5.

Sirovi podaci su sa simulatora kvarova, opisanog u odjeljku 4.1, uz pomoć sustava za prikupljanje podataka, opisanog u odjeljku 4.2, pohranjivani u mjerno računalo. Podaci su putem mrežne infrastrukture dostupni računalnoj infrastrukturi za obradu podataka, na kojoj su

razvijeni modeli dubokog učenja za računalno generiranje značajki prikupljenih podataka zasnovani na konvolucijskoj neuronskoj mreži, pomoću kojih model u konačnici ima mogućnost identifikacije uzroka kvarova u ranoj fazi njihova nastanka.

Konvolucijske neuronske mreže intenzivnu primjenu u području računalnog vida započinj s rezultatima *AlexNet* arhitekture za klasifikaciju slika na *ImageNet* natjecanju [72]. Ipak, razvoj konvolucijskih neuronskih mreža započeo je kasnih 50-ih godina prošloga stoljeća istraživanjima u neuropsihologiji. Hubel i Wiesel opisali su u svom radu [99] način rada vizualnog korteksa mozga, pri čemu su zaključili kako postoje različiti neuroni za prepoznavanje različitih vrsta složenosti promatranog objekta u primarnom vizualnom korteksu. Pritom procesiranje uvijek započinje sa jednostavnim strukturama, a ne promatranjem objekta u cjelini. Ovo otkriće smatra se temeljem funkcionalnosti dubokih struktura konvolucijske neuronske mreže, jer upravo početni konvolucijski slojevi služe prepoznavanju osnovnog obrisa odnosno značajki ulaznog objekta. Nakon toga, objavljen je niz istraživanja u neuroznanosti, a 1980. godine Fukushima u svom radu [100] predlaže računalnu samorganizirajuću arhitekturu za učenje značajki ulaznih objekata, inspiriranu istraživanjima Hubela i Wiesela, sastavljenu od nekoliko konvolucijskih slojeva sa receptivnim poljima koje sadrže vektore težina i nazvanu *Neocognitron*. *Neocognitron* se smatra prvom dubokom arhitekturom umjetne neuronske mreže.

Nekoliko godina kasnije, Yann LeCun [85] primjenjuje algoritam povratnog prostiranja pogreške na Fukushiminu arhitekturu i omogućuje efikasno nadgledano učenje takvog tipa mreža. Može se reći kako su konvolucijske neuronske mreže prisutne od početka 80-ih godina prošloga stoljeća, ali njihova pojačana primjena događa se upravo sada, a potiču je velika količina dostupnih podataka, mogućnosti modernih računala, kao i sposobnost obrade velike količine podataka na grafičkim procesorima.

Ipak, kako je detaljnije objašnjeno u poglavlju 3, većina dosadašnje primjene dubokih konvolucijskih neuronskih mreža vezana je uz područje računalnog vida. U posljednje vrijeme moguće je pronaći istraživanja manjeg opsega vezana uz ograničenu primjenu konvolucijskih neuronskih mreža i u drugim područjima, od kojih su neka i u području održavanja.

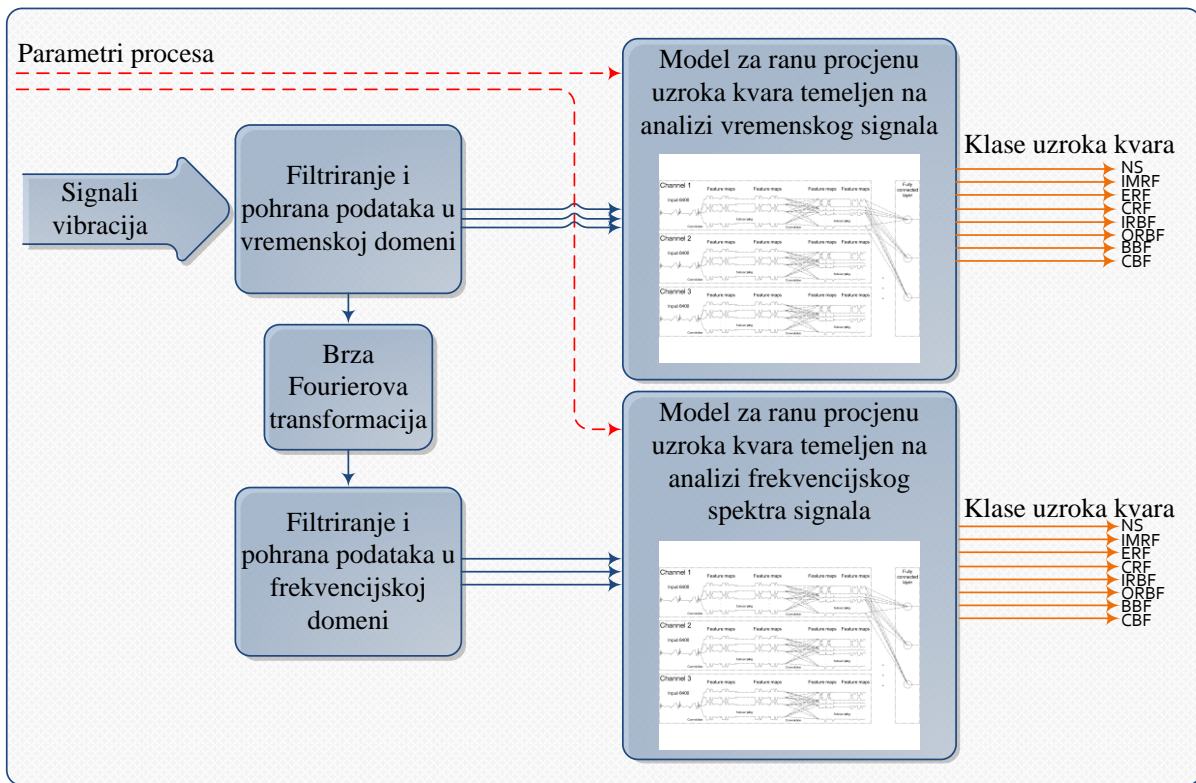
Temeljni razlozi, zbog kojih je moguće tvrditi da CNN može pronaći svoju primjenu u održavanju rotacijske opreme proizlaze iz opisa takvih mreža.

- CNN generalno pokazuju robusnost pri transformaciji i promjeni ulaznih podataka. Ova karakteristika ključna je u području obrade vibracijskog signala, zato što se podaci sa

senzora vibracija mogu značajno razlikovati (različito vrijeme snimanja, različita duljina uzorka signala, šum u vremenskom signalu), bez da se njihove temeljne značajke i karakteristike promjene. Također, neke vrste značajki se ponavljaju na različitim mjestima te je potrebno obratiti pozornost na njihovo značenje i pravilno ih prepoznati pri obradi.

- Autonomnost izdvajanja značajki iz izvornih datoteka je druga ključna karakteristika CNN-a, koja u teoriji omogućuje da ulazni signal u mrežu bude sirovi signal (signal sa senzora zvuka, vibracija ...). Prethodno (odjeljak 2.4) su definirani nedostaci konvencionalnih pristupa inteligentnom održavanju po stanju u vidu potrebe za ekspertom u definiranju i izdvajanju značajki. U ovom slučaju neće biti potrebno ručno izdvajanje značajki koje oduzima mnogo vremena te iziskuje detaljno poznavanje domene koja se obrađuje. Konvolucijske neuronske mreže u tom kontekstu omogućuju reprezentativno učenje, tj. učenje složenih reprezentacija iz izvornog, sirovog signala.

Kako je prikazano na slici 6.2, procjena uzroka kvara tijekom rada rotacijske opreme realizirana je uz pomoć 2 modela, koji će detaljnije biti opisani u narednim odjeljcima. Navedeni modeli, kako je prethodno spomenuto, temeljeni su na konvolucijskim neuronskim mrežama s mogućnošću automatskog učenja značajki. Snimljeni signali pohranjeni su u vremenskoj i frekvencijskoj domeni, a potom su primjenjeni za razvoj modela. Važno je naglasiti, kako snimljeni signali u vremenskoj domeni ne sadržavaju nikakve dodatne značajke, već se u pohranjenim datotekama nalaze isključivo snimljeni signal u vremenu. Signal u frekvencijskoj domeni je predobrađeni signal tj. radi se o značajkama koje su izdvojene brzom Fourierovom transformacijom. Razlog razvoja ovog modela u sklopu disertacije temelji se na provjeri sposobnosti duboke konvolucijske neuronske mreže da uči na temelju značajki već prethodno obrađenog signala. Također, dodatni razlog jest činjenica kako pojedini sustavi za prikupljanje podataka standardno prikupljaju podatke u frekvencijskoj domeni, dok podaci u vremenskoj domeni nisu dostupni.



Slika 6.2 Blok shema procesa razvoja modela za ranu procjenu uzroka kvara rotacijske opreme

6.1 Priprema podataka za učenje i testiranje modela

Zadatak duboke konvolucijske neuronske mreže u ovom slučaju može se opisati kao klasifikacijski zadatak u kojem naučeni model, na temelju ulaznih podataka procjenjuje klasu trenutnog stanja opreme, koja je u procesu učenja definirana kao ciljna varijabla.

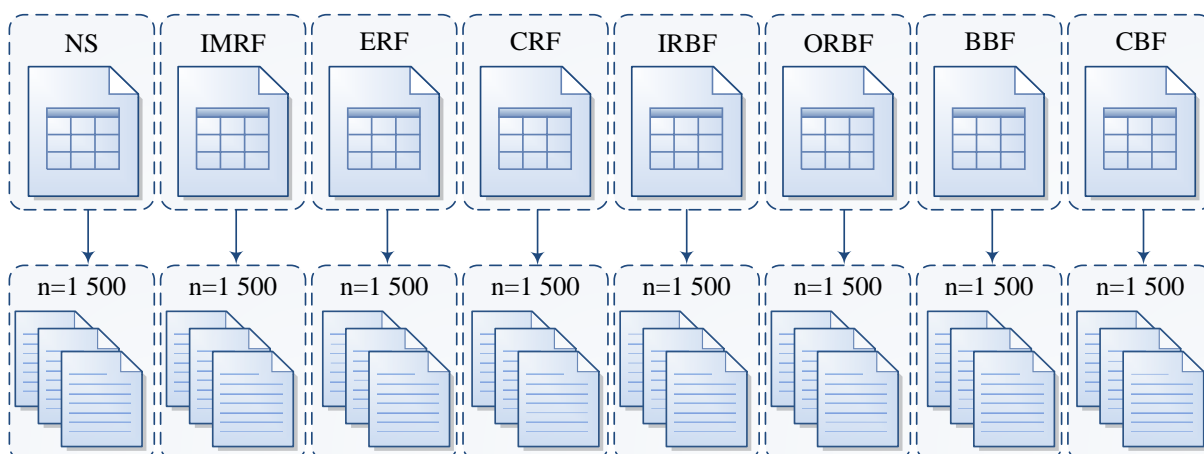
Tijekom provođenja eksperimenta, prikupljeni su podaci za 8 različitih simuliranih stanja rotacijske opreme, na način da su pojedini nadogradni moduli (opisani u odjeljku 5.1) prije ispitivanja montirani na simulator kvarova. U tablici 6.1 su definirani korišteni nadogradni moduli za simuliranje stanja opreme.

Tablica 6.1 Korišteni nadogradni moduli za simulaciju stanja opreme

Simulirano stanje opreme	Korišteni nadogradni moduli
Normalno stanje (NS)	/
Neravnoteža rotora (IMBF)	Rotor sa dodatnom masom osi okomitoj na os vrtnje
Ekscentričnost rotora (ERF)	Ekscentričan rotor
Nagnutost rotora (CRF)	Nagnuti rotor

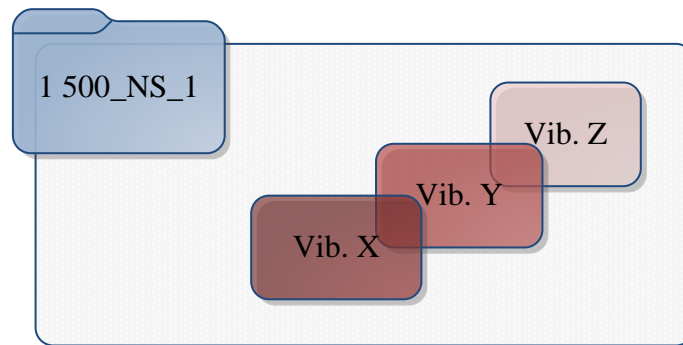
Simulirano stanje opreme	Korišteni nadogradni moduli
Oštećenje vanjske staze kotrljanja kotrljajućeg ležaja (ORBF)	Kuglični ležaj <i>ER12K</i> sa predoštećenom vanjskom stazom kotrljanja
Oštećenje unutarnje staze kotrljanja kotrljajućeg ležaja (IRBF)	Kuglični ležaj <i>ER12K</i> sa predoštećenom unutarnjom stazom kotrljanja
Oštećenje kotrljajućeg elementa kotrljajućeg ležaja (BBF)	Kuglični ležaj <i>ER12K</i> sa predoštećenim kotrljajućim elementom
Kombinirano oštećenje kotrljajućeg ležaja (CBF)	Kuglični ležaj <i>ER12K</i> sa predoštećenjem na unutarnjoj stazi kotrljanja, vanjskoj stazi kotrljanja i kotrljajućem elementu

Podaci prikupljeni i pohranjeni tijekom izvođenja eksperimenta u vremenskoj domeni pripremljeni su na način da svako mjerenje predstavlja jednu datoteku, pa tako definirana podjela znači da svaka klasa snimljenog stanja sadrži 1 500 datoteka, kako je prikazano na slici 6.3.



Slika 6.3 Podjela podataka unutar svake klase uzroka kvara

Svaka datoteka s podacima pojedinog mjerenja unutar svake pojedine klase snimljenog stanja u nazivu sadrži podatak o učestalosti vrtnje i klasi trenutnog stanja (prema tablici 5.1) te podatak o broju mjerenja odnosno ponovljenog pokusa ($n = 1, 2, 3 \dots 1\,500$), kako je prikazano na slici 6.4.

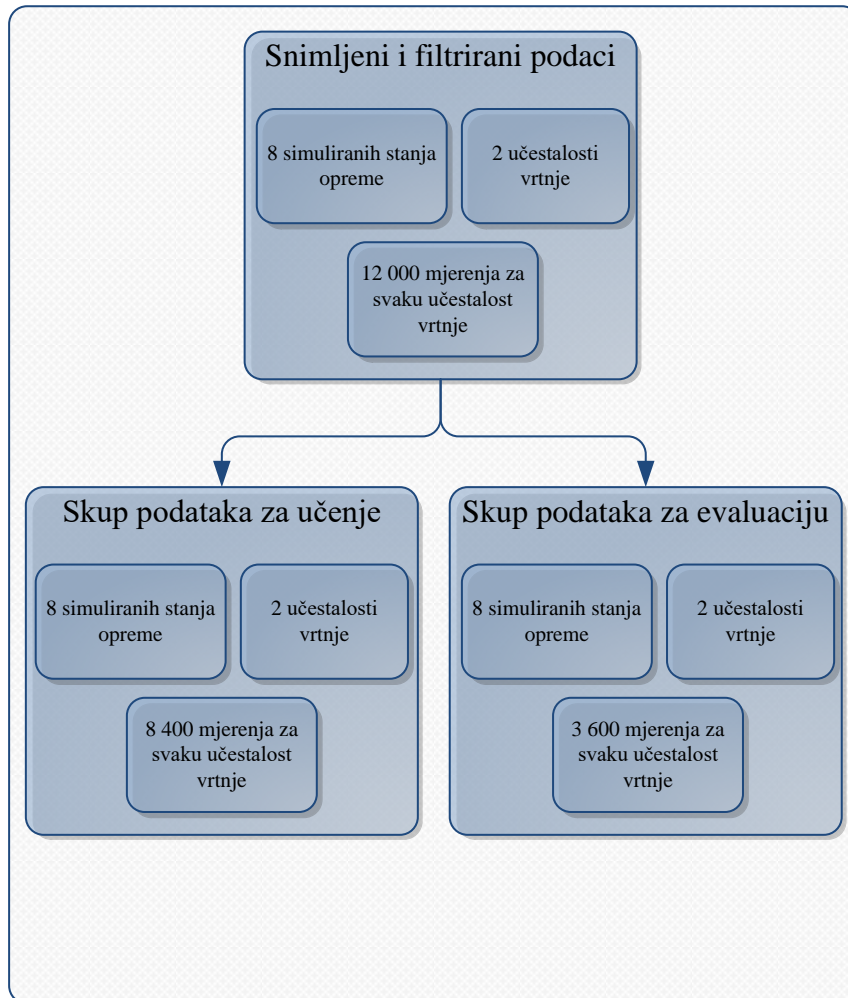


Slika 6.4 Sadržaj pojedine datoteke ispitnog mjerenja

Umjetne neuronske mreže za nadgledano učenje dizajnirane su s ciljem mogućnosti procjene izvan skupa za učenje, pri čemu je važno održati generalizacijska svojstva mreže koja se definiraju kao sposobnost procjene na skupu podataka različitom od skupa na kojem je učila.

Prema definiranom planu eksperimenta (opisanog u odjeljku 5.2), prikupljeno je ukupno 12 000 mjerenja na konstantnoj učestalosti vrtnje od 1 000 o/min te 12 000 mjerenja na konstantnoj učestalosti vrtnje od 1 500 o/min.

Za potrebe učenja i evaluacije modela, prikupljeni skup podataka podijeljen je u podskupine podataka na način da je nasumičnim odabirom 70 % od ukupnog zbroja podataka za konstantnu učestalost vrtnje od 1 500 o/min označeno kao skup podataka za učenje, dok će preostalih 30 % podataka označenih kao testni skup podataka služiti za evaluaciju modela, čime će se osigurati sposobnost generalizacije umjetne neuronske mreže.



Slika 6.5 Blok shema skupova podataka za učenje i testiranje

Skup podataka za učenje sastoji se od 16 800 mjerenja nastalih ponavljanjem mjerenja za 8 simuliranih stanja opreme uz 2 učestalosti vrtnje, dok se unutar skupa podataka za evaluaciju nalazi stohastički određenih 7 200 mjerenja, izdvojenih iz ukupne sume svih snimljenih i filtriranih podataka (vidljivo na slici 6.5).

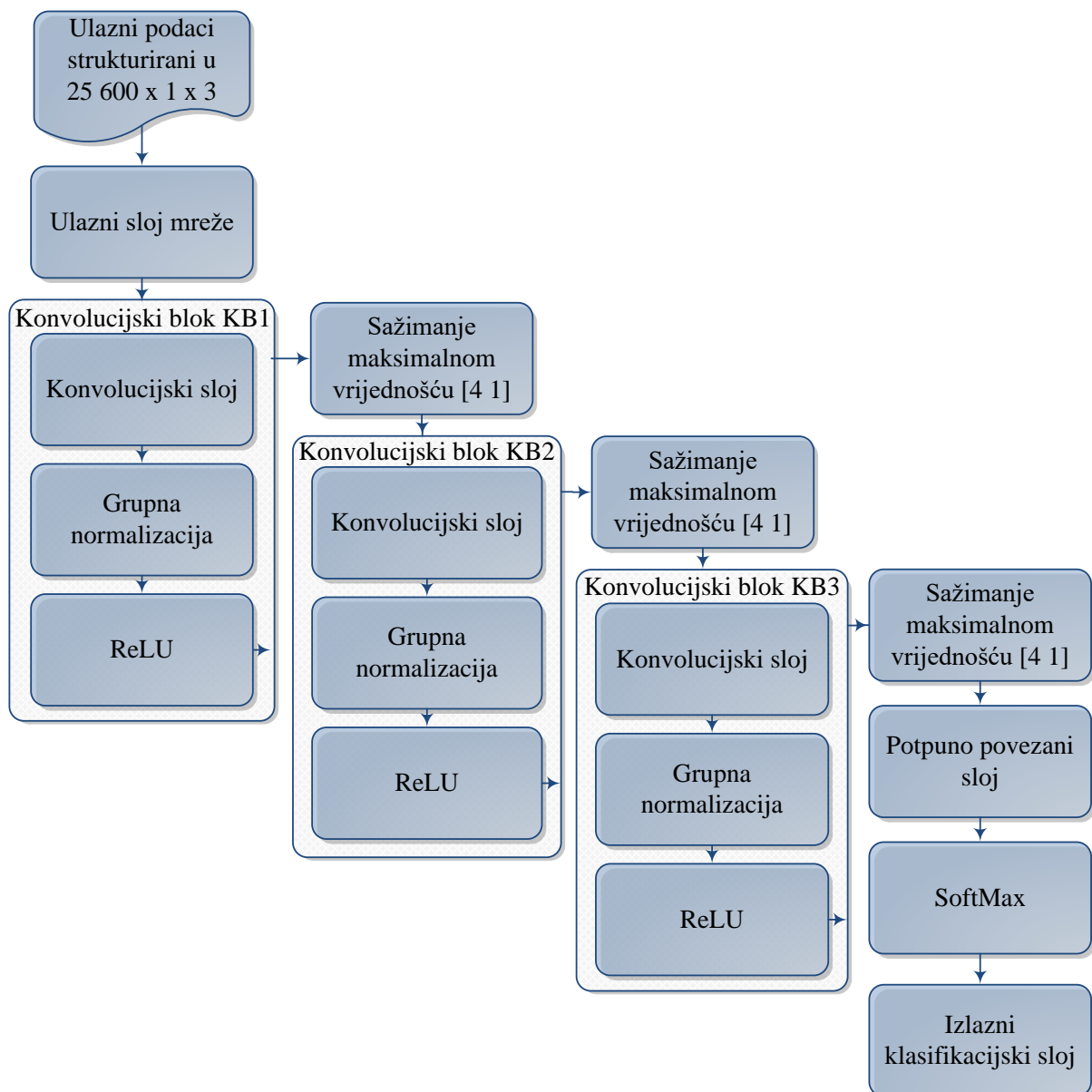
Opisana podjela skupova podataka vrijedi za skupove podataka u vremenskoj, kao i za podatke pohranjene u frekvencijskoj domeni.

6.2 Struktura modela temeljenog na konvolucijskoj neuronskoj mreži

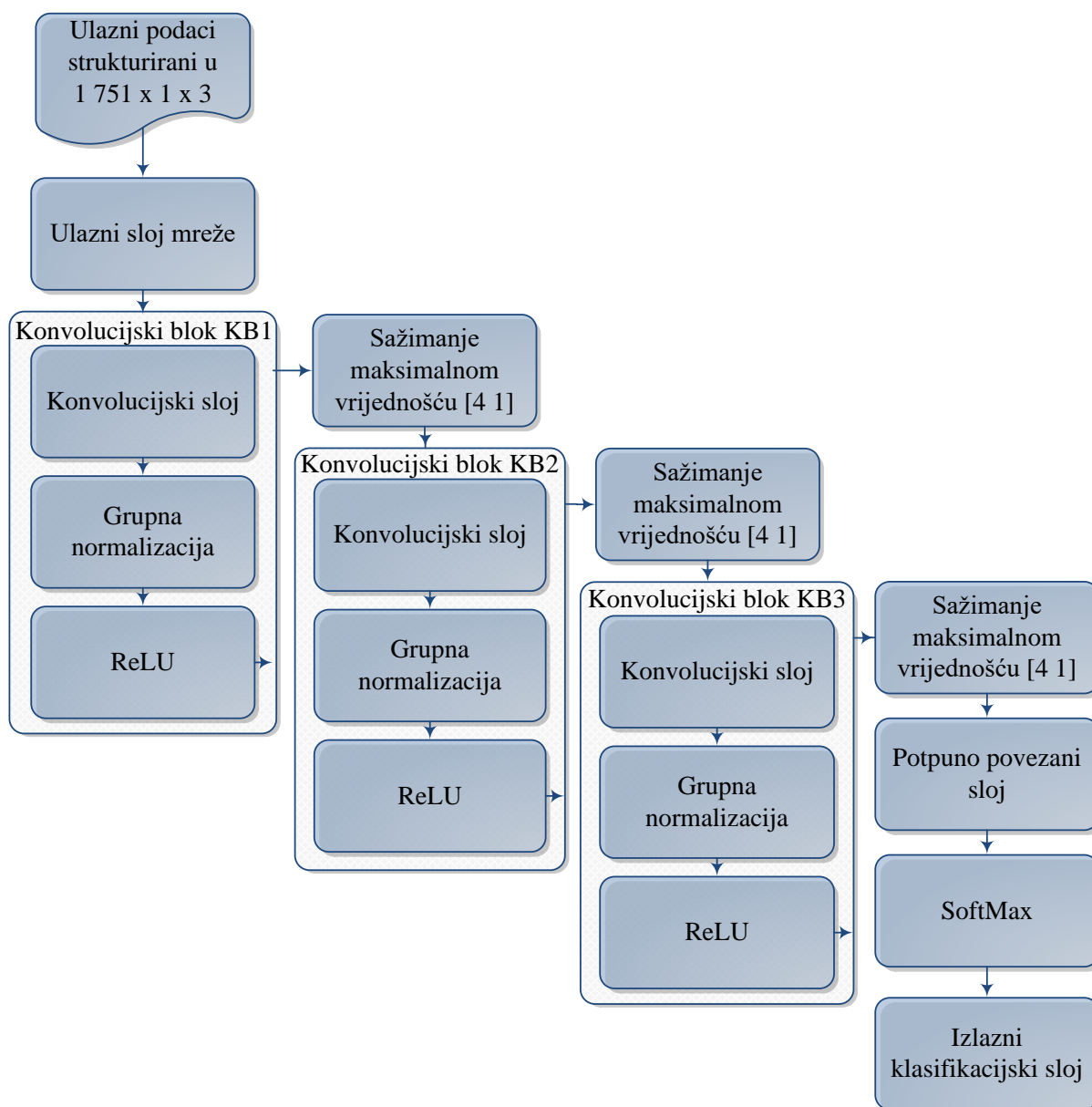
Razvijeni modeli rane procjene uzroka kvara temelje se na primjeni modela duboke konvolucijske neuronske mreže sa ulaznim signalima u vremenskoj i frekvencijskom domeni, kako prikazuje slika 6.2.

Konvolucijska neuronska mreža detaljno je opisana u poglavlju 3. Generalno, može se reći kako je svaka duboka konvolucijska neuronska mreža sastavljena od ulaznog sloja, naslaganih slojeva u kojima se izmjenjuju konvolucijski sloj, aktivacijski sloj (ReLU, vidi odjeljak 3.1) i sloj sažimanja te potpuno povezanog sloja sa završnom aktivacijskom funkcijom i slojem izlaznih (klasnih) neurona čija aktivacija predstavlja pripadnost određenoj klasi.

Na slici 6.6 prikazan je inicijalni dizajn modela temeljenog na dubokoj konvolucijskoj neuronskoj mreži sa ulaznim signalom u vremenskoj domeni. Radi lakšeg snalaženja, za ovaj model koristit će se skraćeni naziv MVD.



Slika 6.6 Blok shema inicijalnog modela temeljenog na konvolucijskoj neuronskoj mreži sa ulaznim signalom u vremenskoj domeni (MVD)



Slika 6.7 Blok shema inicijalnog modela temeljenog na konvolucijskoj neuronskoj mreži sa ulaznim signalom u frekvencijskoj domeni (MFD)

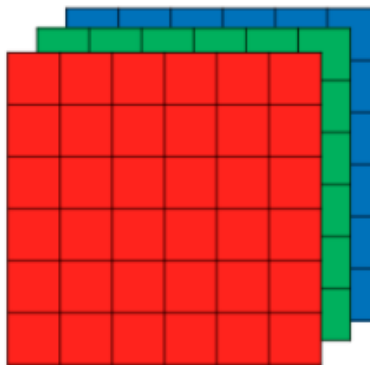
Slika 6.7 prikazuje inicijalni model temeljen na konvolucijskoj neuronskoj mreži sa ulaznim signalom u frekvencijskoj domeni. Radi lakšeg snalaženja, za ovaj model koristit će se skraćeni naziv MFD. Dizajn inicijalnog modela sa ulaznim signalom u frekvencijskoj domeni razlikuje se od modela sa ulaznim signalom u vremenskoj domeni samo u veličini ulaznog signala. Konačni cilj računalnog sustava svodi se na točnu procjenu tipa uzroka kvara, pri čemu model temeljen na signalima opisanim u frekvencijskoj domeni kao ulazni signal zaprima već predobrađene i transformirane signale, tj. signale vremenske domene koji su podvrgnuti brzom

Fourierovoj transformaciji. U idealnom slučaju, kvalitetna procjena tipa uzroka kvara modelom temeljenim na podacima u vremenskoj domeni bit će dodatno potvrđena procjenom modela s podacima iz frekvencijske domene. Stoga će se u narednim odjeljcima detaljno opisati model temeljen na konvolucijskoj neuronskoj mreži s ulaznim signalom u vremenskoj domeni, a dodatno će biti navedene razlike takvog modela i modela sa ulaznim signalom u frekvencijskoj domeni.

U narednim odjeljcima bit će detaljnije opisani korišteni slojevi modela konvolucijske neuronske mreže.

6.2.1 Ulazni sloj mreže

Prethodno je navedeno kako je razvoj konvolucijskih neuronskih mreža potaknut njihovim mogućnostima primjene u području računalnog vida. Upravo stoga ulazni sloj konvolucijske neuronske mreže očekuje sliku kao ulaznu matricu. Slika kao ulazni objekt najčešće je definirana kao trodimenzionalna matrica piksela slike u kojoj se na jednoj osi nalazi duljina, a na drugoj širina slike, dok se po dubini matrice slažu slojevi R, G i B spektra, kako je prikazano na slici 6.8.



Slika 6.8 Prezentacija RGB ulaznog signala mreže

Prema podacima prikupljenim eksperimentom, vibracijski signal troosnog akcelerometra sadrži mjerenja u osima X, Y i Z, a jedno mjerenje (uzorak) definirano eksperimentom sadrži 25 600 zapisa.

Upravo se pri otkrivanju mogućnosti da se rezultati mjerenja troosnih akcelerometara prikazuju kao matrice dimenzija *broj_zapisa_mjerenja* x *broj_akcelerometara* x *broj_kanala_akcelerometra* inicijalno stvorila ideja o primjeni konvolucijske neuronske mreže za rješavanje klasifikacije uzroka kvarova koja je opisana u ovom istraživanju.

Ulazni sloj mreže definiran je dimenzijama ulaznog signala. Ulazni signal u slučaju MVD modela predstavlja jedno mjerenje (uzorak) jednog akcelerometra te je ukupnih dimenzija 25 600 x 1 x 3 (tablica 6.2), dok je u slučaju MFD modela ukupnih dimenzija 1 750 x 1 x 3 (tablica 6.3)

Tablica 6.2 Dimenzije ulaznog signala MVD modela

Broj podataka uzorka	25 600
Broj korištenih akcelerometara	1
Broj osi akcelerometra	3

Tablica 6.3 Dimenzije ulaznog signala MFD modela

Broj podataka uzorka	1 750
Broj korištenih akcelerometara	1
Broj osi akcelerometra	3

Dodatno, nad ulaznim slojem podataka provodi se *zero-center* normalizacija, čime se postiže približna jednakost skale vrijednosti za pojedine dimenzije. To se postiže dijeljenjem svake dimenzije (kanala) sa standardnom devijacijom tako da je centrirana na nulu. Centriranje na vrijednost nule znači oduzimanje srednje vrijednosti iz svake od tih dimenzija tako da je *podatkovni oblak* svake dimenzije usmjeren prema ishodištu.

Normalizacija je potrebna iz razloga što se tijekom procesa učenja mreže početne ulazne vrijednosti signala množe (uz pomoć težina) i dodaju (uz pomoć pomaka) kako bi se uzrokovale aktivacije koje se potom vraćaju s gradijentima.

6.2.2 Konvolucijski blokovi mreže

Konvolucijski skriveni slojevi duboke neuronske mreže zaduženi su za automatsko učenje značajki ulaznog signala definiranog u ulaznom sloju mreže. Slojevi funkcioniraju tako da pomoću jezgri pretvaraju ulazne signale u izlazne, tj. u tzv. mape značajki. Proces se sastoji od toga da se na ulazni signal primijeni unaprijed određeni broj jezgri, na svaku se jezgru doda pomak te na kraju nelinearna aktivacijska funkcija kako bi se dobila izlazna matrica (mapa značajki), kako je detaljnije objašnjeno u odjeljku 3.1.

Svaki konvolucijski sloj slijedi i sloj normalizacije grupe (engl. *batch normalization*), i uvođenje nelinearnosti preko aktivacijske funkcije (ReLU).

U prethodnom odjeljku objašnjeno je na koji način je izvedena normalizacija ulaznog signala podataka za učenje modela s ciljem ubrzavanja učenja. S obzirom na činjenicu kako je za proces učenja normalizacija ulaznog sloja korisna, slično je moguće primijeniti i za vrijednosti u skrivenim slojevima. Normalizacija grupe provodi se s ciljem smanjenja iznosa pomaka skrivenih slojeva (kovarijantnog pomaka), a prema postupku objašnjenom u odjeljku 3.5.

Na temelju literature i provedenih inicijalnih ispitivanja, definirana su tri tipa konvolucijskih blokova (*KB1*, *KB2*, *KB3*).

6.2.3 Slojevi sažimanja

Slojevi sažimanja razvijenog modela nalaze se iza konvolucijskih blokova. Razlog korištenja slojeva sažimanja u kontekstu razvoja ovog modela leži u sposobnosti slojeva sažimanja da modelu namaknu otpornost odnosno invarijantnost na translacije naučenih značajki unutar ulaznog signala [81]. Konkretno, u ulaznom vibracijskom signalu to znači sposobnost otkrivanja vibracijskih značajki i u slučaju da su one uslijed vanjskog utjecaja pomaknute u odnosu na naučene vrijednosti. Nadalje, sažimanje smanjuje broj parametara koji se mogu učiti i posljedično smanjuje mogućnost efekta pretreniranja. Prema prikazanom na slici 6.6 vidljivo je kako se iza konvolucijskih blokova tipa *KB1*, *KB2* i *KB3* nalazi sloj sažimanja maksimalnom vrijednošću. Fizikalno, sažimanje se može promatrati kao svojevrsan oblik smanjivanja vrijednosti koeficijenta uzorkovanja (engl. *downsampling*) uz zadržavanje naučenih karakteristika iz mapa značajki. Zadržavanje maksimalnih vrijednosti donekle osigurava prijenos značajnih informacija, iako pritom ne otklanja u potpunosti rizik od gubitaka.

6.2.4 Potpuno povezani sloj i *SoftMax*

Kao što je prethodno definirano, konvolucijski slojevi modela djeluju na način da otkrivaju i izdvajaju značajke ulaznog vibracijskog signala. Postupkom sažimanja, posljednji konvolucijski blok sažima se maksimalnom vrijednošću. Mape značajki nakon sažimanja spajaju se u jedan skup podataka te se potom postupkom poravnavanja (engl. *flattening*) transformiraju u 1D skup podataka. Takvi podaci tada djeluju kao ulaz u neuron(e) potpuno povezanog sloja koji izvodi skalarni umnožak ulaznih podataka i zadane težine uz dodavanje pomaka, kako bi izračunao vrijednost izlaza (jedan izlaz po neuronu).

Prema tome, može se zaključiti kako će potpuno povezani sloj mreže obavljati isti zadatak, kao i u slučaju klasične unaprijedne neuronske mreže, no sam ulaz će biti različit s obzirom na prethodne konvolucijske blokove i slojeve sažimanja. Neuroni u potpuno povezanom sloju

tijekom postupka treninga opisuju ulazne podatke uz pomoć redefiniranja težina i pomaka, omogućavajući ispravnu kategorizaciju klase u narednom *SoftMax* sloju.

6.3 Učenje i optimizacija modela

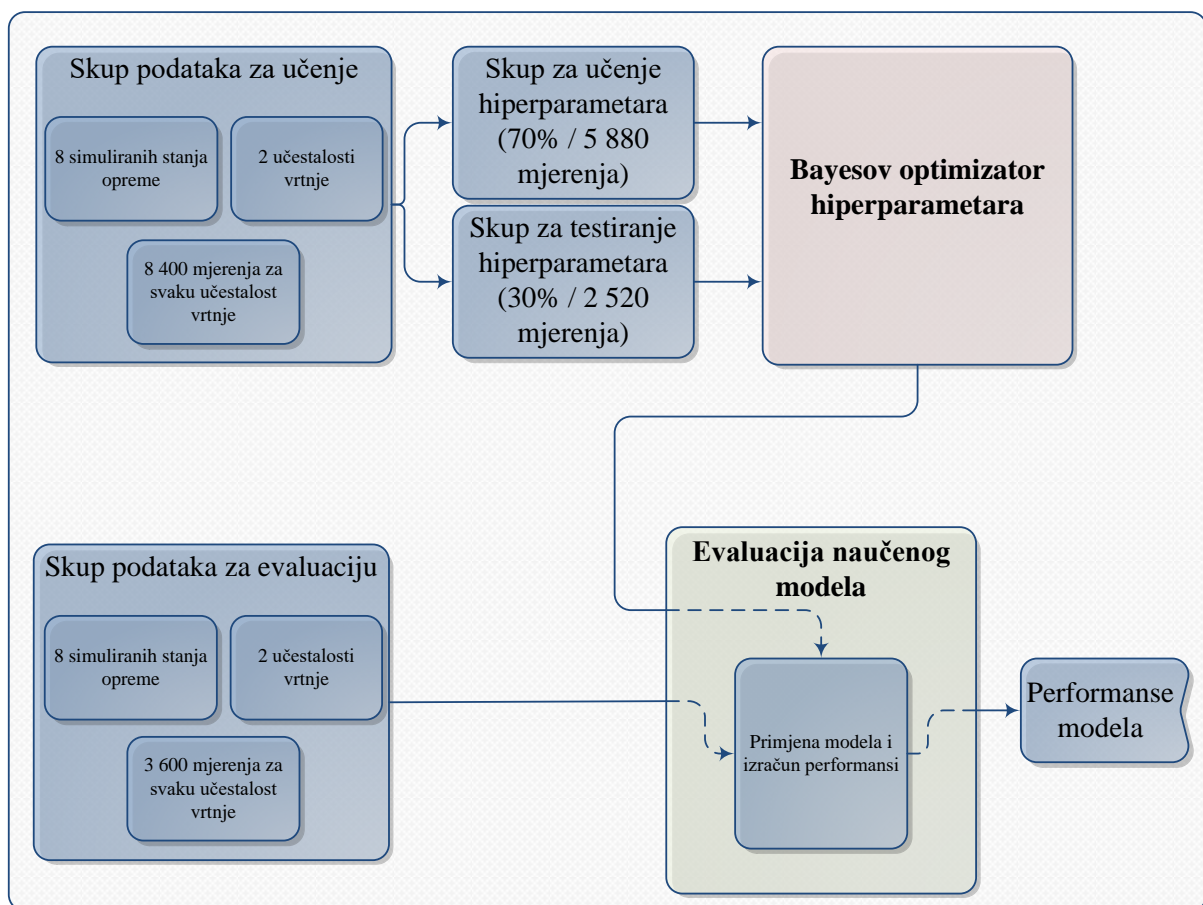
Cilj modela za ranu procjenu uzroka kvarova rotacijske opreme svodi se na kvalitetno preslikavanje ulaznih na željene izlazne vrijednosti. Pritom navedeni cilj mora biti ispunjen kako u fazi učenja, tako i u fazi naknadnog korištenja mreže. Konvolucijske neuronske mreže spadaju u kategoriju nadziranog strojnog učenja ili učenja s učiteljem (engl. *supervised machine learning*). Način učenja modela s učiteljem podrazumijeva osiguran skup uzoraka sa informacijom o njihovoj pripadnosti nekoj od klasifikacijskih skupina. Učenje modela temeljenog na umjetnim konvolucijskim neuronskim mrežama izvedeno je na način da se postignu dobra generalizacijska svojstva, tj. točan odziv na bilo koju kombinaciju ulaznih vrijednosti iz područja učenja. U tom je smislu važno osigurati dovoljan broj adekvatnih uzoraka koji opisuju prostor učenja, a odabir konačnog modela temeljiti na rezultatima ostvarenima u fazi evaluiranja, tj. na uzorcima koji nisu korišteni u učenju. Osim na temelju rezultata dobivenih u fazi evaluiranja mreže, važan je utjecaj hiperparametara na konačan model. Najvažniji hiperparametri opisani su u odjeljku 3.7. Umjetne konvolucijske neuronske mreže sadrže veći broj hiperparametara čija promjena utječe na ishod učenja te je prilikom postupka učenja potrebno pronaći skup hiperparametara koji daje najbolji ishod, tj. najbolje performanse modela na testnim skupovima podataka.

Generalno, nastoji se kreirati model s mogućnošću prepoznavanja uzorka u ulaznom vibracijskom signalu, no ne postoji standardno definirana metoda odabira ispravnih hiperparametara. Intuitivno, povećanje dimenzija hiperparametra *veličine jezgre* znači i obuhvaćanje većeg broja susjednih vrijednosti signala u kojem se nastoji pronaći karakteristični uzorak, dok veći *broj jezgri* znači i veći broj reprezentacija odnosno veći broj mapa značajki. Povećavanje bilo kojeg od prethodno navedenih parametara povećava kompleksnost računanja i vrijeme učenja. Kako je navedeno u odjeljku 3.8, ugađanje parametara često se obavlja jednostavnim strategijama poput manualnog ili mrežnog pretraživanja. Međutim, manualno traženje ili mrežno pretraživanje se vrlo teško prilagođava povećanju broja hiperparametara (dimenzionalnosti) i općenito ne funkcionira dobro za više od nekoliko parametara. Alternativne tehnike globalne optimizacije, kao što su genetski algoritmi su fleksibilnije, ali također obično zahtijevaju više procjena nego što ih je vremenski moguće izvesti, posebno u prisutnosti nesigurnosti. Bayesova optimizacija započinje izgradnjom jednostavnog surogatnog modela ciljne funkcije korištenjem Gaussovog procesa na temelju opažanja dostupnih iz prethodnih eksperimenata odnosno prethodnih iteracija. Predviđanja i procjene nesigurnosti se

kombiniraju kako bi se dobila funkcija stjecanja, koja kvantificira vrijednost trenutnog opažanja određene točke. Optimizira se funkcija stjecanja kako bi se pronašla najbolja točka, a zatim se nakon promatranja ishoda postavi novi surogatni model i proces se ponavlja do konvergencije. Cijeli proces je prilagodljiv u smislu da se predviđanja i procjene nesigurnosti ažuriraju u skladu s novim promatranjima.

Dakle, kako bi se kreirao optimalan model, potrebno je razviti i kreirati računalom upravljani proces, koji će, iskorištavajući dostupne podatke iz skupa za učenje (vidi sliku 6.3), Bayesovom optimizacijom iznaći optimalan skup hiperparametara te na taj način omogućiti najbolje moguće rezultate naučenog modela na skupu za učenje.

Slika 6.9 u koracima definira proces učenja modela rane procjene tipa uzroka kvara rotacijske opreme. Početni skup podataka podijeljen je na skup podataka za učenje i evaluiranje modela. Nadalje, skup podataka za učenje, s obzirom da je to ulazni skup u algoritam strojnog učenja hiperparametara duboke konvolucijske neuronske mreže, u drugom se koraku dodatno dijeli u omjeru 70 : 30 % u korist skupa za učenje hiperparametara.



Slika 6.9 Blok shema procesa učenja modela rane procjene tipa uzroka kvarova primjenom Bayesovog optimizatora

Ovi podaci ulaze u proces Bayesove optimizacije (za detalje vidi poglavlje 3.8.4), koja se provodi pod uvjetima objašnjenima u narednim poglavljima (6.3.1 i 6.3.2).

Nakon izvršavanja petlje algoritma Bayesovog optimizatora, model za procjenu tipa uzroka kvara temeljen na dubokoj konvolucijskoj neuronskoj mreži i naučen uz pomoć optimalne kombinacije hiperparametara testira se uz pomoć skupa podataka za evaluiranje te se izračunavaju njegove performanse.

Bayesova optimizacija je pristup optimizaciji računalno intenzivnih funkcija cilja. Pristup se pokazao prikladnim za optimizaciju kontinuiranih domena sa manjim brojem dimenzija i tolerantan je na stohastički šum u evaluaciji funkcija. Bayesov optimizator kreira računalno povoljnu surogatnu funkciju cilja i kvantificira nesigurnost procjene surogatne funkcije koristeći Bayesovu tehniku strojnog učenja, tj. Gaussov slučajni proces. Zatim koristi funkciju stjecanja definiranu od tog surogata da bi odlučila u kojem području domene će se istraživati.

Glavni razlog korištenja Bayesove optimizacije u kontekstu iznalaženja vrijednosti hiperparametara modela za procjenu tipa uzroka kvarova temeljenu na konvolucijskoj neuronskoj mreži je mogućnost korištenja podataka stvarne funkcije cilja kako bi se konstruirala procjena funkcije poznata kao surogat. Naime, poznavajući proces učenja konvolucijske neuronske mreže, zaključuje se kako modeliranje istog predstavlja računalno intenzivni proces. Surogat funkcija Bayesovog optimizatora može se koristiti kao zamjena za stvarnu funkciju i brže ispitati kako bi se odabrao kandidat (skup vrijednosti hiperparametara) visoke kvalitete i visokog stupnja sigurnosti procjene za sljedeći eksperiment.

Intuitivno, nesigurnost procjene surogat modela će biti najveća u područjima koja još nisu istražena kroz iteracije. Takva pretpostavka omogućuje da se kreira način za postupno otkrivanje područja istraživanja i pretraživanja funkcije gdje postoji vrlo malo informacija uz iskorištavanje funkcije u područjima za koja se *vjeruje* da daju bolje vrijednosti. Može se reći kako je ovaj kompromis presudan u smislu pokušaja da se globalno minimizira funkcija, bez zapinjanja u lokalnom minimumu, odnosno suboptimalnom rješenju. Funkcija korisnosti koja opisuje način korištenja kompromisa naziva se, kako je objašnjeno u poglavlju 3.8.4, stjecajnom funkcijom.

6.3.1 Stjecajna funkcija Bayesovog optimizatora

Nakon što je model posteriora naučen sa trenutnim skupom zapažanja, sljedeći korak je upotreba naučenog modela za odabir sljedeće iteracije, koja će se pokrenuti. Cilj Bayesove optimizacije je pronaći globalno optimalno rješenje u što je moguće manjem broju vrednovanja

funkcija odnosno broja iteracija. Da bi to bilo moguće, koristi se stjecajna funkcija $a(x)$, koja mjeri predviđenu kvalitetu kandidatskih iteracija. Pravilan odabir između dostupnih stjecajnih funkcija (opisanih u odjeljku 3.8.4) najčešće ovisi o dostupnom vremenskom okviru. U slučaju da je važno pronalaženje globalno optimalnog rješenja, a postoji i potrebno vrijeme za pronalaženje istog, onda bi se moglo provesti iscrpno pretraživanje. Međutim, optimizacijsko vrijeme je ograničeno, te se u nekom trenutku ne dopuštaju daljnje iteracije, već se mora predložiti najbolja pretpostavka. Kako raste broj iteracija, to može postati računalno intenzivno, jer je potencijalno potrebno predvidjeti sve ishode svih načina odabira različitih točaka. Stoga, najčešće korištene funkcije stjecanja su kratkog dometa u smislu što pretpostavljaju da je dopušteno iterirati još jednu točku prije završetka pretraživanja (čak i ako to nije nužno točno). Takav način osigurava ravnotežu između učinkovitosti računanja i uspješnosti optimizacije.

U ovoj tezi, koristi se funkcija *očekivanog poboljšanja* (engl. *expected improvement, EI*), koja, za razliku od funkcije *vjerojatnosti poboljšanja*, nije toliko *pohlepna* te ima manju vjerojatnost zapinjanja u lokalnom minimumu iz razloga što se ne fokusira isključivo na odgovor na pitanje hoće li biti poboljšanja u novoj točki, već i na iznos poboljšanja.

Prema [89], ova je funkcija primjer učinkovite funkcije. Ipak, ponekad vrijeme za procjenu točaka funkcije cilja može ovisiti o regiji koja se ispituje. U tom slučaju, Bayesov optimizator ostvaruje bolje rezultate ukoliko se vrijeme izvršavanja ponderira unutar stjecajne funkcije. Tijekom procjene funkcije cilja, optimizator izračunava dodatni Bayesov model ovisnosti vremena evaluacije funkcije cilja o trenutnoj poziciji x , pa konačna stjecajna funkcija glasi:

$$EI_{pS}(x) = \frac{EI_Q(x)}{\mu_S(x)}, \quad (6.1)$$

gdje je $\mu_S(x)$ aritmetička sredina posteriora vremenskog modela Gaussovog procesa. Dodatno, kako bi se izbjeglo zapinjanje u lokalnom minimumu, koristi se modifikacija koja omogućuje promjenu ukoliko algoritam procjeni predugo zadržavanje unutar istog prostora pretraživanja. Ukoliko je $\sigma_F(x)$ standardna devijacija funkcije cilja posteriora u točki x , σ standardna devijacija dodanog šuma, tada vrijedi:

$$\sigma_Q^2(x) = \sigma_F^2(x) + \sigma^2. \quad (6.2)$$

U ovom slučaju, nakon svake iteracije Bayesov optimizator izračunava zadovoljava li sljedeća točka x uvjet $\sigma_F(x) < t_\sigma \sigma$, gdje je t_σ pozitivni broj. Ukoliko je uvjet zadovoljen, pretpostavlja

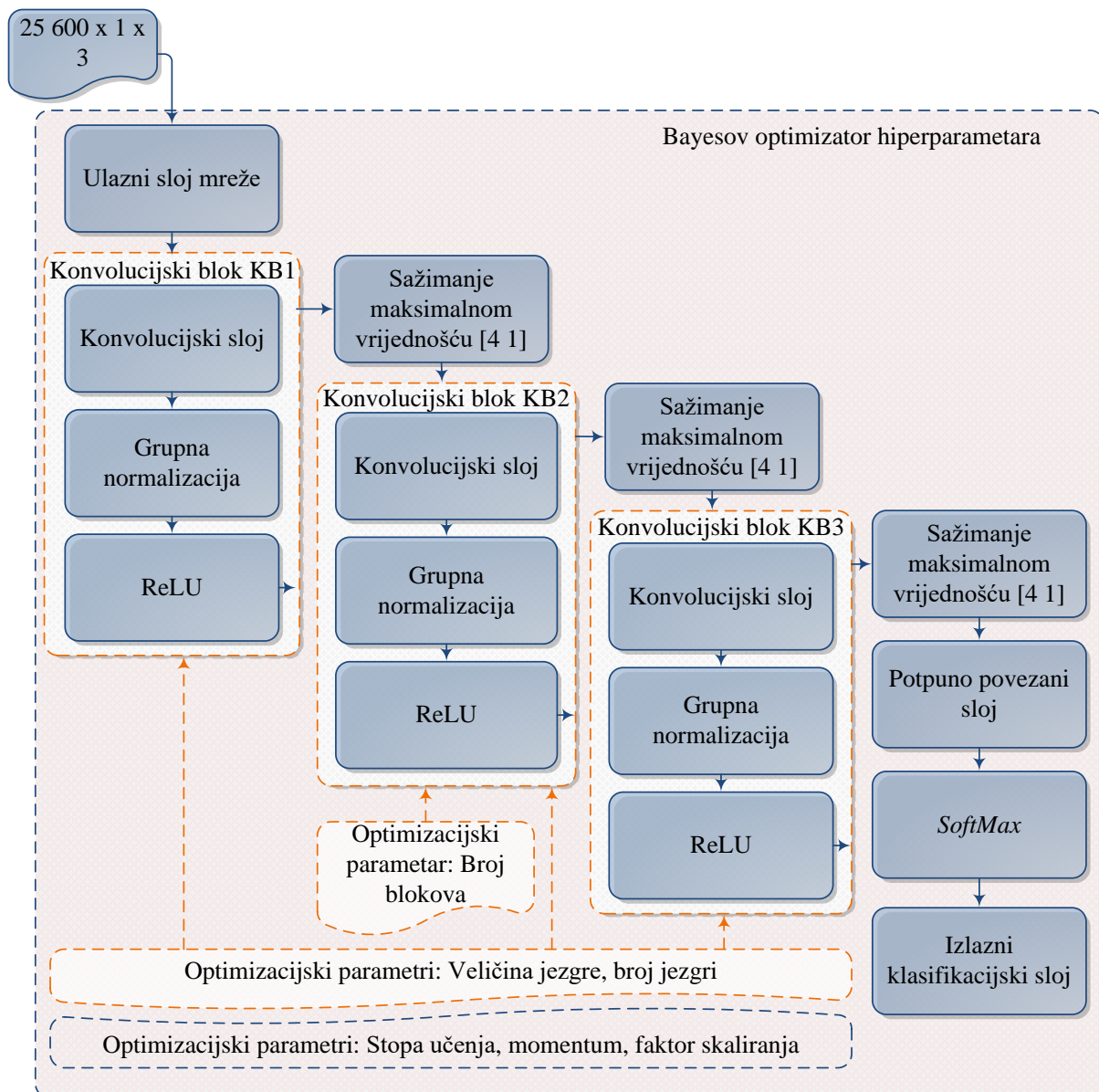
se da je potrebna promjena prostora za točku x , te funkcija stjecanja modificira svoju kernel funkciju (vidi odjeljak 3.8.4) multipliciranjem σ sa brojem iteracija [101]. Ova modifikacija podiže vrijednost varijance σ_Q za točke promatranja. Potom se generiraju nove točke na temelju nove kernel funkcije. Ukoliko je i prostor nove točke x prekomjerno istražen, funkcija množi σ s dodatnim faktorom 10 i do maksimalno 5 puta pokušava na isti način pronaći točku x u prostoru koji nije prekomjerno istražen. Može se zaključiti kako je na ovaj način, uz pomoć t_σ , moguće kontrolirati odnos između istraživanja novih točaka za bolje globalno rješenje i koncentriranje uz točke koje su već ispitane.

Za potrebe ovog rada koristit će se Bayesov optimizator sa stjecajnom funkcijom *očekivanog poboljšanja* s uključenim dodatkom koji omogućuje određivanje vremena trajanja izračuna funkcije cilja, kao i mogućnost promjene prostora za točku x ukoliko se procjeni predugo iskorištavanje trenutnog prostora istraživanja.

Dakle, Bayesov optimizator će maksimizirati funkciju stjecanja na način da u prvoj fazi procjenjuje najmanju moguću srednju vrijednost posteriorne funkcije $\mu_Q(x_{best})$ uzorkovanjem nekoliko tisuća točaka unutar granica varijabli, uzimajući nekoliko najboljih mogućih točaka (niska srednja vrijednost) i poboljšavajući ih koristeći lokalno pretraživanje, kako bi pronašli prividno najbolju moguću točku. Pritom vrijednost funkcije stjecaja ovisi o modeliranoj distribuciji posteriora, a ne o uzorku funkcije cilja, te se može brzo izračunati.

6.3.2 Odabir hiperparametara za optimizaciju

Na slici 6.10 prikazana je inicijalna struktura mreže pripremljene za fazu optimizacije hiperparametara Bayesovom optimizacijom.



Slika 6.10 Blok shema pripreme za učenje optimizacijom parametara

Za model rane procjene uzroka kvarova rotacijske opreme, definirane su tri grupe hiperparametara, koje će biti optimizirane primjenom Bayesove optimizacije:

1. hiperparametri na razini mreže:
 - a. koeficijent učenja η ,
 - b. momentum,
 - c. faktor skaliranja;
2. hiperparametri na razini konvolucijskih blokova:
 - a. veličina jezgri,
 - b. broj jezgri, i
3. hiperparametar na razini konvolucijskog bloka KB2:

a. broj naslaganih blokova / dubina.

Bayesov optimizator izvršava proces optimizacije hiperparametara, pri čemu je po definiciji potrebno, kao ulazne varijable, definirati granice pojedinih hiperparametara. Naime, iako mogućnost optimizacije računalno intenzivnih funkcija koje se ponašaju kao *black-box* čine Bayesov optimizator primjenjivim u većini slučajeva, potrebno je uslijed postojanja ograničenog vremena izvršavanja ograničiti prostor mogućih rješenja. Nizom preliminarnih eksperimenata potvrđen je koncept inicijalno predstavljen u [64], prema kojem je moguć kvalitetniji odziv modela u slučaju manjeg broja većih jezgri unutar prvog od konvolucijskih blokova. Stoga je osim standardnih hiperparametara samog modela konvolucijske neuronske mreže, unutar optimizacijskih varijabli Bayesovog optimizatora dodana varijabla faktora skaliranja. Ova optimizacijska varijabla omogućuje skaliranje hiperparametara konvolucijskih blokova i kreiranje odnosa u veličinama između veličina jezgri i broja jezgri u pojedinim slojevima. Na temelju literature i inicijalnih pokusa određeni rasponi optimizacijskih varijabli prikazani su u tablici 6.4, i time su definirane optimizacijske varijable korištene kao vrijednosti hiperparametara pri provođenju Bayesove optimizacije.

Tablica 6.4 Rasponi optimizacijskih varijabli Bayesovog optimizatora

Optimizacijska varijabla	Raspon	Tip podatka
<i>vKoeff_učenja (vKU)</i>	0,005 – 0,01	Kontinuirani, <i>logaritamska skala</i>
<i>vMomentum (vM)</i>	0,85 – 0,95	Kontinuirani, <i>logaritamska skala</i>
<i>vVeličina_jezgre (vVJ)</i>	4 – 64	Cjelobrojni
<i>vBroj_jezgri (vBJ)</i>	4 – 24	Cjelobrojni
<i>vDubina_sloja (vDS)</i>	0 – 5	Cjelobrojni
<i>vFaktor_skaliranja (vFS)</i>	1 – 3	Cjelobrojni

Tablica 6.5 opisuje optimizacijske varijable korištene kao vrijednosti hiperparametara pri provođenju Bayesove optimizacije. Korištenje varijabli za podešavanje broja jezgri i veličine

jezgri na razini konvolucijskih blokova u kombinaciji sa varijablom faktora skaliranja u konačnici smanjuje ukupan broj optimizacijskih varijabli, tj. smanjuje vrijeme izračuna Bayesovog optimizatora.

Tablica 6.5 Izračun hiperparametara preko optimizacijskih varijabli

Razina hiperparametra	Hiperparametar	Korištene optimizacijske varijable
Mreža	Koef. učenja	v_{KU}
Mreža	Momentum	v_M
KB1	Veličina jezgre	$v_{VJ} \cdot v_{FS}$
	Broj jezgri	v_{BJ}
KB2	Veličina jezgre	v_{VJ} / v_{FS}
	Broj jezgri	$v_{BJ} \cdot v_{FS}$
	Broj naslaganih blokova	v_{DS}
KB3	Broj jezgri	v_{BJ}

Osim optimizacijskih varijabli, Bayesov optimizator podrazumijeva definiranje ulaznih hiperparametra samog algoritma za optimizaciju. Hiperparametri Bayesovog optimizatora opisani su u odjeljcima 3.8.4 i 6.3.1. Tablica 6.6 dodatno ukratko opisuje korištene hiperparametre algoritma Bayesovog optimizatora i njihove vrijednosti.

Tablica 6.6 Hiperparametri Bayesovog optimizatora

Hiperparametar	Opis	Vrijednost
Stjecajna funkcija	Naziv korištene stjecajne funkcije	<i>Očekivano poboljšanje po sekundi (plus)</i>
Udio istraživanja (engl. <i>exploration ratio</i>)	Kontrola udjela istraživanja novog prostora u prostoru rješenja i trenutno dostupnih rezultata	0,5

Hiperparametar	Opis	Vrijednost
Broj početnih točaka	Broj slučajno odabranih točaka za početnu procjenu funkcije cilja	10
Maksimalni broj evaluacija	Ograničenje brojem izvršavanja algoritma	100
Maksimalno vrijeme	Vremensko ograničenje izvršavanja algoritma	120 [h]

6.4 Metrika za evaluaciju modela

Evaluacija modela rane procjene tipa uzroka kvarova u MVD i MFD modelima izvedena je uz pomoć izračunavanja uspješnosti klasifikacije (6.3).

$$UK = \frac{\text{Ispravno klasificirani uzorci}}{\text{Ukupan broj uzoraka}} \cdot 100 [\%] \quad (6.3)$$

Iako najčešće korištena u praktičnoj primjeni i stoga vrlo često nametnuta kao prvi izbor pri klasifikaciji rezultata, uspješnost klasifikacije odabrana je zbog činjenice o postojanju jednakog broja uzoraka za svaku klasu pri kojoj je uspješnost klasifikacije dokazana kao povoljna evaluacijska mjera.

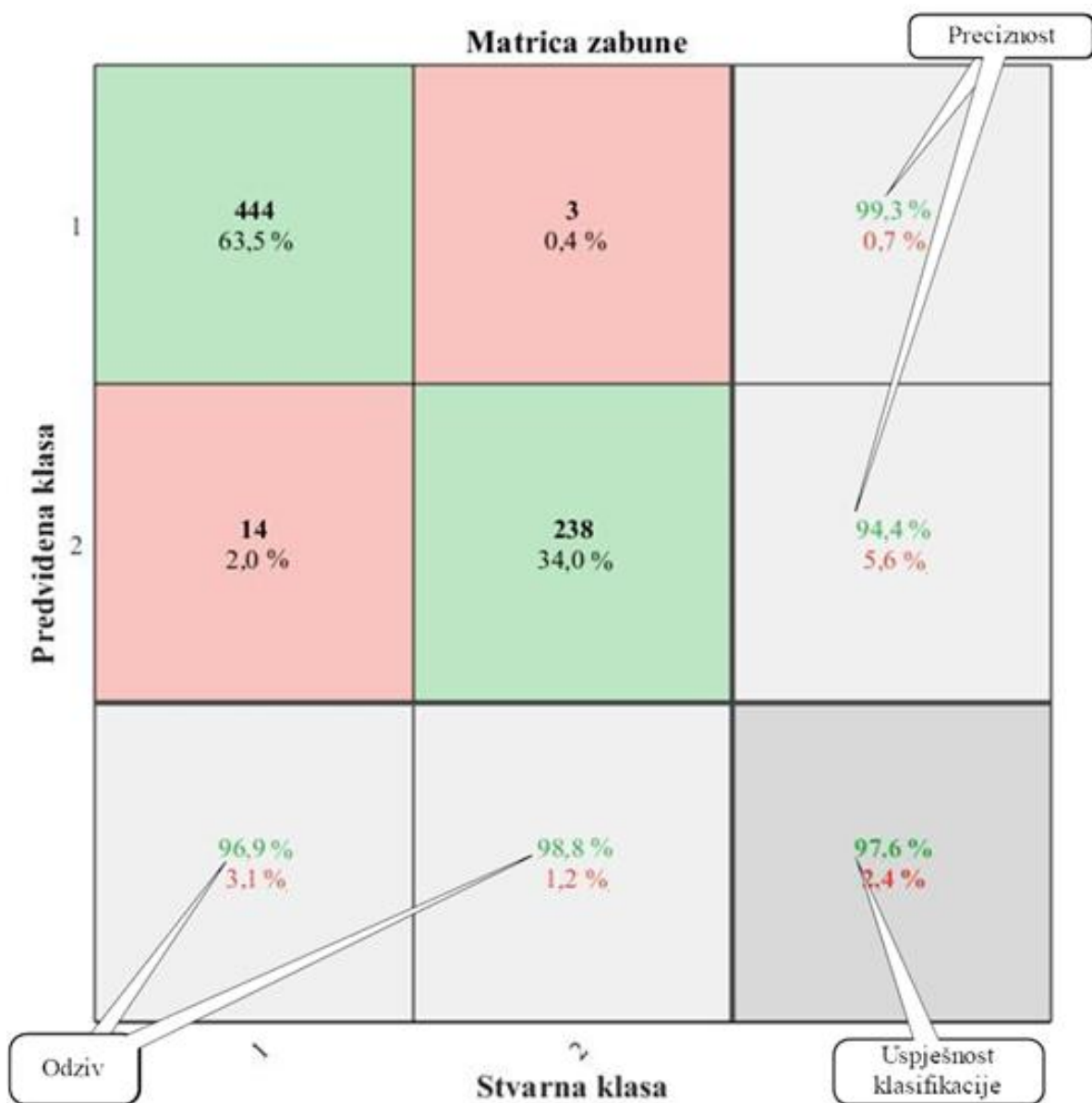
Dakle, uspješnost klasifikacije primijenjena je u ovoj tezi u dva slučaja kao:

- evaluacijska mjera u koraku optimizacije Bayesovim optimizatorom i kao
- evaluacijska mjera u koraku konačne evaluacije na testnom skupu podataka

Osim uspješnosti klasifikacije, koristit će se i mjera matrice zabune (engl. *confusion matrix*) ili matrice pogreške, koja u matričnom obliku sumira i prezentira rezultate predikcije klasifikacijskog modela. Slika 6.11 prikazuje primjer matrice zabune. Za razliku od mjere uspješnosti klasifikacije, matrica zabune prikazuje detaljniji uvid u načine na koji klasifikacijski model griješi, tj. opisuje tipove pogrešaka kojima rezultira podvrgavanje testnog skupa klasifikacijskom modelu. Uobičajeno, matrica zabune prezentirana je na način da svaki redak matrice predstavlja rezultate predviđanja za odgovarajuću klasu u tom retku, dok svaki stupac predstavlja stvarnu klasu. Čelije na dijagonali lijevo-gore → desno-dolje definiraju broj i postotak ispravnih klasifikacija od strane treniranog klasifikatora, dok čelije izvan dijagonale predstavljaju pogrešno klasificirana predviđanja. Zadnji sivi stupac predstavlja preciznost klasifikatora (engl. *precision*) pri predviđanju svake klase, dok zadnji sivi redak predstavlja

postotak odziva (engl. *recall*) klasifikatora unutar svake klase. Drugim riječima, preciznost opisuje koliko često je predviđanje ispravno ukoliko se predvidi određena klasa, dok odziv mjeri koliko se ispravno predviđa određena stvarna klasa.

Visoka vrijednost odziva i niska preciznost određene klase tumači se kao pristranost klasifikatora određenoj klasi, dok je visoka vrijednost preciznosti i niska vrijednost odziva znak svojevrsne *konzervativnosti* modela. Naime, pri takvoj situaciji model ispravno pogađa klasu kad se odluči za tu klasu, no niska vrijednost odziva istovremeno znači da će se rijetko odlučiti za tu klasu. U idealnom slučaju, nastoji se trenirati klasifikator s visokom vrijednošću odziva i preciznosti.



Slika 6.11 Primjer matrice zabune

Vrijednosti preciznosti i odziva za prvi redak i prvi stupac na prethodnom primjeru moguće je izračunati pomoću sljedećih formula:

$$\begin{aligned} \text{preciznost}_1 &= \frac{T_1}{T_1 + F_1} \\ \text{odziv}_1 &= \frac{T_1}{T_1 + F_2}, \end{aligned} \tag{6.4}$$

gdje je T_1 broj ispravno predviđenih stvarnih *klasa 1*, F_1 broj neispravno predviđenih *klasa 1*, a F_2 broj pogrešno predviđenih primjera stvarne *klase 1*.

6.5 Rezultati i evaluacija modela

Na temelju opisanog procesa učenja modela uz optimizaciju hiperparametara (6.3) te metrike definirane u odjeljku 6.4, proveden je postupak kreiranja modela. Kao ulazni podaci u proces učenja korišteni su trening podaci definirani kao ulazni skup u odjeljku 6.1. Provedeni postupak učenja mreže Bayesovom optimizacijom uključuje ponavljanje procesa učenja u 100 koraka s ciljem pronalaska najbolje kombinacije hiperparametara, odnosno hiperparametara koji će za dani slučaj kreirati model temeljen na konvolucijskoj neuronskoj mreži s najboljim odzivom na testnom skupu podataka. Kako je već spomenuto u odjeljku 3.8.4, Bayesova optimizacija definirana je kao proces izrade računalno nezahtjevnog surogatnog modela funkcije cilja te njegove upotrebe za odabir najboljih hiperparametara, koji se potom evaluiraju na (računalno zahtjevnoj) stvarnoj funkciji cilja kroz proces evaluiranja naučene mreže. U narednim odjeljcima, dani su i komentirani rezultati kreiranih optimiziranih modela. Modeli su tijekom provođenja Bayesove optimizacije testirani na skupu od 2 880 mjerenja (vidi sliku 6.9), a odabrani modeli dalje su testirani na testnim skupovima.

6.5.1 Model procjene kvarova na temelju signala vremenske domene

Učenje modela temeljenog na vremenskoj domeni signala podrazumijeva, kako je opisano u odjeljku 6.2, ulazni signal u obliku *vibracijske slike* dimenzije $25\ 600 \times 1 \times 3$. Model je učen na način da su u tijeku učenja optimizirani hiperparametri primjenom Bayesove optimizacije. Tablica 6.7 ispisuje trag korištenih vrijednosti varijabli za Bayesovu optimizaciju, pri čemu je u posljednjoj desnoj koloni upisana i pohranjena vrijednost uspješnosti klasifikacije za model sa hiperparametrima određenim uz pomoć vrijednosti zadanih varijabli (tablica 6.5 sadrži poveznice između korištenih varijabli optimizatora i hiperparametara modela).

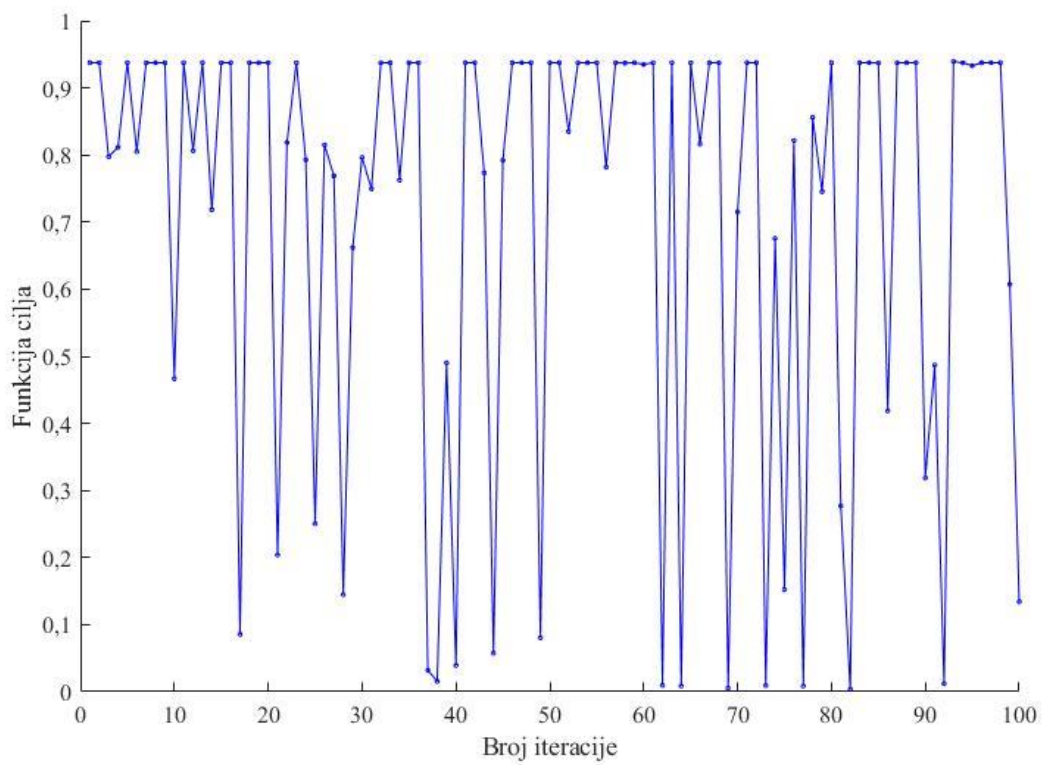
Tablica 6.7 Iteracije Bayesove optimizacije MVD

Iter. br.	Parametri bloka			Parametri mreže		vFS	UK, [%]
	vDS	vVJ	vBJ	vKU	vM		
	3	19	22	0,009537615	0,930210499	2	6,25
2	5	20	23	0,005005383	0,948109861	2	6,25
3	7	48	6	0,007292659	0,885674324	1	20,26
4	1	4	9	0,005329222	0,904420751	3	18,85
5	6	19	9	0,005983309	0,893040747	1	6,25
6	5	46	12	0,007862605	0,894010526	3	19,46
7	6	30	17	0,009967388	0,920641252	1	6,25
8	0	63	15	0,008803733	0,925468492	2	6,25
9	7	58	14	0,009455051	0,860467085	1	6,25
10	1	58	16	0,008794614	0,854849168	2	53,33
11	1	56	15	0,006160614	0,899588245	1	6,25
12	8	59	24	0,005608477	0,865869602	2	19,37
13	4	50	18	0,007170948	0,916896791	2	6,25
14	4	28	21	0,008546834	0,874225102	1	28,15
15	7	36	15	0,0081729	0,934402286	1	6,25
16	4	25	21	0,009900929	0,854211681	2	6,25
17	6	9	7	0,006510735	0,850259787	3	91,43
18	7	6	17	0,008834908	0,947294575	2	6,25
19	3	23	8	0,005338734	0,863292887	2	6,25
20	6	52	7	0,007506564	0,949598131	1	6,25
21	8	5	20	0,007052035	0,855378839	3	79,60
22	5	20	19	0,009854558	0,852419102	3	18,13
23	3	30	9	0,006745864	0,851017101	3	6,25
24	4	7	7	0,009478775	0,850131175	3	20,69
25	6	7	5	0,006239116	0,854324896	2	74,94
26	6	9	19	0,007279008	0,852826678	3	18,51
27	6	15	8	0,007100455	0,850599681	3	23,12
28	7	4	13	0,006464241	0,859212549	2	85,50
29	8	4	9	0,006676175	0,852356247	1	33,83
30	6	8	11	0,0063921	0,866008991	3	20,40
31	7	5	8	0,006411501	0,850922195	2	25,04
32	1	29	16	0,00567529	0,920040224	1	6,25
33	0	45	6	0,009439555	0,93775271	1	6,25
34	0	4	8	0,009682252	0,913881154	1	23,77
35	1	64	19	0,005098241	0,943761658	1	6,25
36	0	34	11	0,007530798	0,949321941	1	6,25
37	3	47	23	0,005863081	0,851303631	2	96,79

<i>Iter. br.</i>	<i>Parametri bloka</i>			<i>Parametri mreže</i>			
	<i>vDS</i>	<i>vVJ</i>	<i>vBJ</i>	<i>vKU</i>	<i>vM</i>	<i>vFS</i>	<i>UK, [%]</i>
38	3	49	4	0,008086732	0,850186035	1	98,45
39	1	48	13	0,007091635	0,851069708	1	50,99
40	3	49	11	0,005935973	0,854004309	1	96,07
41	3	50	4	0,007531115	0,855637551	3	6,25
42	3	48	14	0,009050892	0,850502114	1	6,25
43	1	11	10	0,009914873	0,891634354	2	22,66
44	2	22	6	0,005121531	0,850112767	3	94,25
45	4	28	6	0,006131355	0,86353305	3	20,79
46	6	24	6	0,006710024	0,851611034	3	6,25
47	1	56	5	0,005213882	0,881966092	3	6,25
48	1	54	11	0,005172002	0,932316928	2	6,25
49	2	32	10	0,005079343	0,909436721	1	91,94
50	2	45	10	0,008734928	0,921426782	1	6,25
51	3	32	10	0,005458976	0,943129071	1	6,25
52	0	63	10	0,007545129	0,946245192	1	16,49
53	3	53	14	0,006822839	0,938876051	1	6,25
54	1	48	14	0,00985445	0,902074519	1	6,25
55	5	57	5	0,009614005	0,879984956	3	6,25
56	1	61	6	0,007536699	0,906935563	2	21,83
57	6	20	4	0,009906637	0,868722455	3	6,25
58	6	47	5	0,006487174	0,875748073	3	6,29
59	2	57	13	0,005025328	0,921753325	3	6,25
60	5	41	12	0,007885977	0,93870915	3	6,49
61	4	28	18	0,005000799	0,914043427	1	6,25
62	7	21	9	0,005068711	0,875558518	1	99,05
63	5	28	9	0,00997637	0,949607536	3	6,25
64	2	44	17	0,005228173	0,852550041	1	99,15
65	3	12	9	0,006196088	0,949906362	3	6,25
66	7	24	19	0,005027459	0,896836744	1	18,35
67	3	31	14	0,009998846	0,916891138	3	6,25
68	2	31	11	0,009997742	0,948873169	1	6,25
69	1	54	13	0,007856801	0,863103248	1	99,44
70	0	5	14	0,005103822	0,893075141	1	28,49
71	8	25	6	0,008286463	0,868617146	2	6,25
72	0	45	6	0,009991109	0,868580542	3	6,25
73	8	21	5	0,0054734	0,85325762	3	99,07
74	2	10	5	0,006205519	0,866181494	1	32,44
75	4	8	6	0,008055991	0,857718852	1	84,74
76	7	22	11	0,005012509	0,946633661	3	17,86

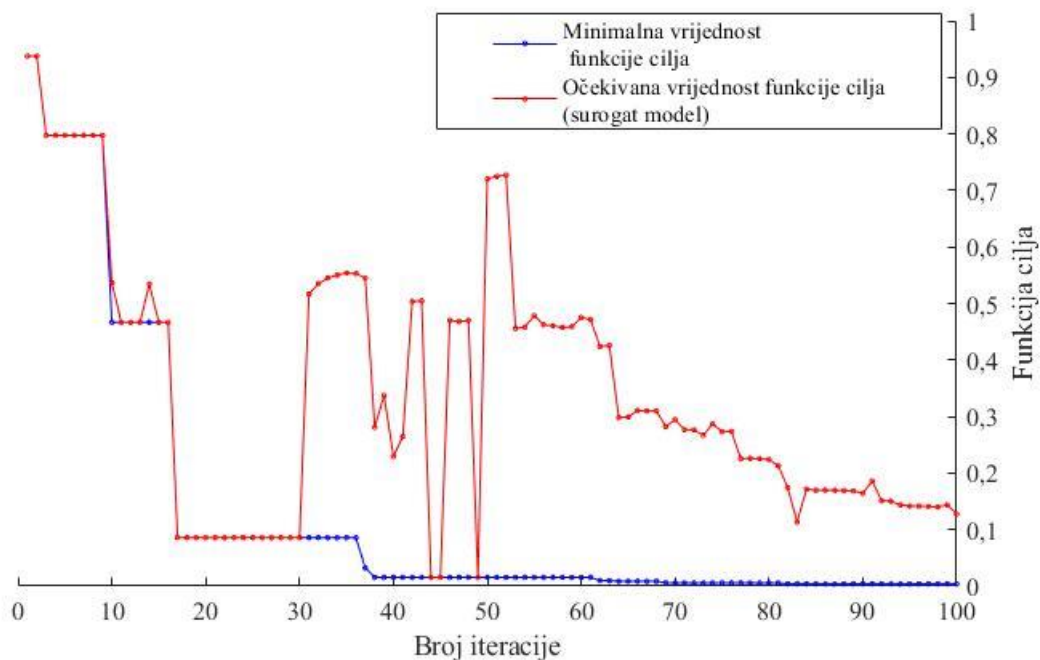
<i>Iter. br.</i>	<i>Parametri bloka</i>			<i>Parametri mreže</i>			
	<i>vDS</i>	<i>vVJ</i>	<i>vBJ</i>	<i>vKU</i>	<i>vM</i>	<i>vFS</i>	<i>UK, [%]</i>
77	2	61	5	0,005104293	0,858719858	1	99,13
78	2	22	16	0,009988759	0,900493584	3	14,40
79	1	63	18	0,005000311	0,909556932	2	25,46
80	2	40	14	0,006244873	0,930714552	3	6,25
81	0	43	4	0,005534201	0,850320097	3	72,30
82	1	23	5	0,005172443	0,850790306	2	99,60
83	8	55	14	0,008630373	0,917767036	3	6,25
84	2	41	4	0,008157641	0,881219103	3	6,25
85	2	47	17	0,008535289	0,899050938	1	6,29
86	2	42	6	0,009973953	0,875898905	1	58,13
87	1	46	16	0,005050972	0,949049183	3	6,25
88	0	38	11	0,008920832	0,949695179	3	6,25
89	0	40	7	0,009777045	0,934508705	3	6,25
90	8	41	5	0,005518029	0,851401292	3	68,13
91	4	51	6	0,006406688	0,895784996	3	51,33
92	1	51	6	0,005037766	0,872358819	1	98,81
93	0	16	19	0,007764337	0,913001226	3	6,07
94	8	9	4	0,005028529	0,893509545	3	6,25
95	8	43	11	0,009930735	0,891247934	1	6,71
96	8	7	5	0,009996169	0,949766832	1	6,25
97	8	53	15	0,008130095	0,949925005	3	6,25
98	0	61	13	0,00640099	0,949137004	2	6,25
99	0	15	20	0,00839511	0,855138008	1	39,27
100	7	27	4	0,005019353	0,870150673	1	86,59

Slika 6.12 prikazuje trend funkcije cilja, pri čemu valja naglasiti kako nativno Bayesov optimizator minimizira funkciju cilja te se unutar Bayesove optimizacije odvija proces koji nastoji smanjiti pogrešku uspješnosti klasifikacije. Moguće je utvrditi kako je najveću vrijednost funkcije cilja od 99,6 % uspješno klasificiranih podataka optimizacijski model pronašao nakon izvršene 82 iteracije.



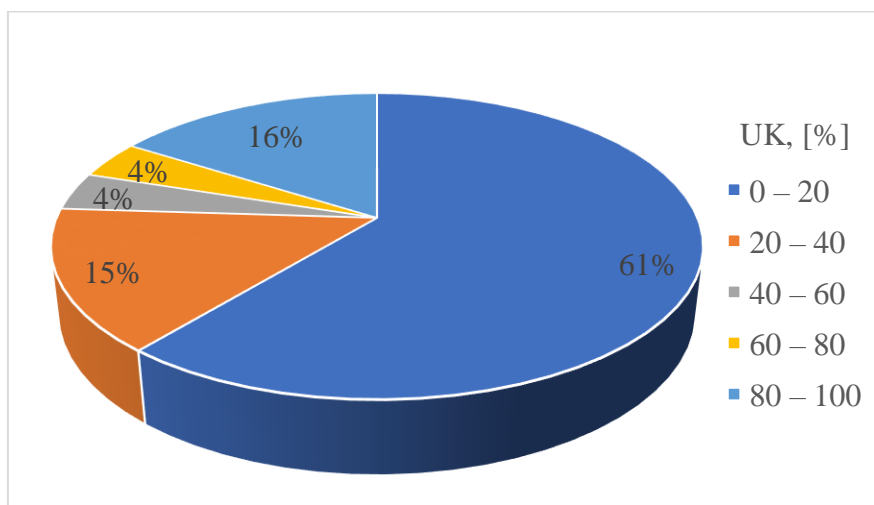
Slika 6.12 Trend funkcije cilja u procesu optimizacije

Dodatno, kreiran je grafikon minimalne i očekivane funkcije cilja (slika 6.13), na temelju kojeg se može utvrditi kako se s porastom broja iteracija surogatni model prilagođava istraženim vrijednostima.



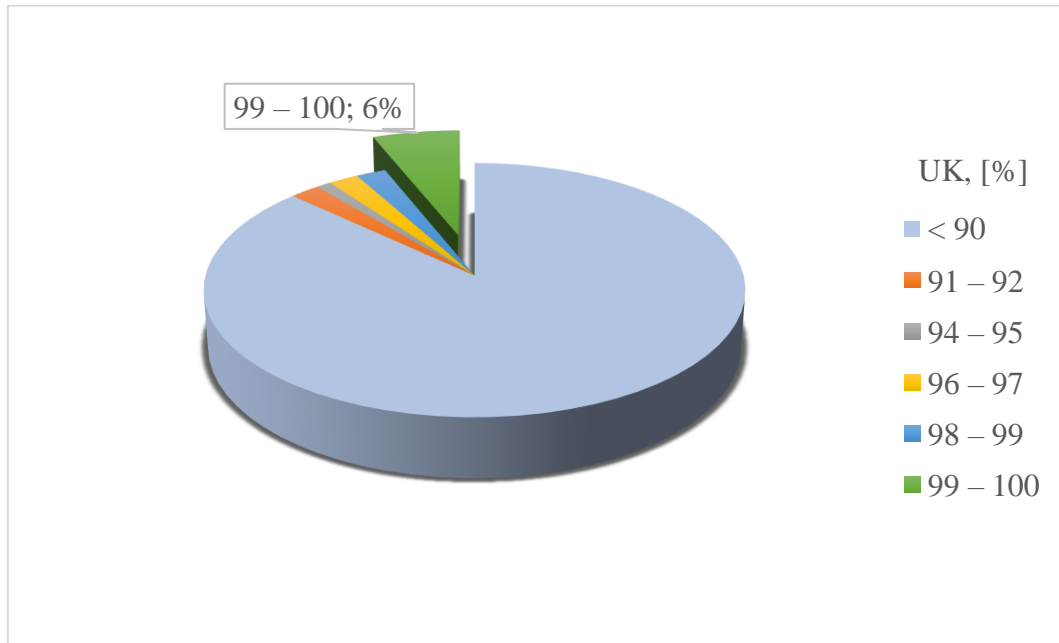
Slika 6.13 Grafikon minimalne i očekivane funkcije cilja tijekom procesa optimizacije

Također, vrlo je indikativna važnost pravilno odabranih vrijednosti hiperparametara (koje bi bez primjene Bayesove optimizacije bilo vrlo teško odrediti) za kreiranje uspješnog modela, s obzirom da je moguće iz tablice izračunati kako čak 75 % rezultata ne prelazi uspješnost klasifikacije od 50 %. Na slici 6.14 prikazana je podjela rezultata uspješnosti klasifikacije na način da su podaci grupirani u pet razreda, a u kojem je jasno vidljivo kako 80 % kreiranih modela ne prelazi uspješnost klasifikacije od 60 %, a više od 60 % modela ne dostiže ni uspješnost klasifikacije od 20 %.



Slika 6.14 Grafikon rezultata uspješnosti klasifikacije UK, [%]

Prema podacima sa slike 6.15, iz kojih je vidljivo kako 6 % od ukupnih rezultata Bayesovog klasifikatora ima uspješnost klasifikacije veću od 99 %, kreirana je tablica 6.8 koja prikazuje rezultate modela za 6 % najboljih iteracija.



Slika 6.15 Grafikon grupiranih rezultata uspješnosti klasifikacije UK, [%]

Vrijednosti varijabli korištenih tijekom Bayesove optimizacije te dobivenih rezultata potvrđuju potrebu za korištenjem tehnike za optimizaciju hiperparametara pri razvoju modela za ranu procjenu uzroka kvara rotacijske opreme primjenom duboke konvolucijske neuronske mreže.

Tablica 6.8 Najboljih 6 % kombinacija vrijednosti varijabli Bayesove optimizacije

Iter. br.	Parametri bloka			Parametri mreže			UK, [%]
	vDS	vVJ	vBJ	vKU	vM	vFS	
82	1	23	5	0,005172443	0,850790306	2	99,60
69	1	54	13	0,007856801	0,863103248	1	99,44
64	2	44	17	0,005228173	0,852550041	1	99,15
77	2	61	5	0,005104293	0,858719858	1	99,13
73	8	21	5	0,0054734	0,85325762	3	99,07
62	7	21	9	0,005068711	0,875558518	1	99,05

Slijedom procesa prikazanog na slici 6.9, mrežu sa najboljim rezultatom uspješnosti klasifikacije unutar Bayesovog optimizatora provjerava se na testnom skupu podataka. Uspješnost klasifikacije na testnom skupu podataka iznosi također 99,60 % odnosno 29

pogrešno klasificiranih testnih uzoraka na ukupno 7 200 testnih mjerenja. Visoki rezultat na skupu za učenje popraćen jednako uspješnim rezultatom klasifikacije testnih primjera može biti protumačen kvalitetnim modelom sa neizraženim svojstvima podnaučenosti ili prenaučivosti.

U nastavku, slika 6.16 prikazuje matricu zabune za evaluirani model duboke konvolucijske neuronske mreže sa najboljim rezultatom na skupu za testiranje hiperparametara (*iteracija br. 82*).

Predviđena klasa	1000BBF	437 6.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.6% 1.4%		
	1000CBF	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%	
	1000CRF	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1000ERF	0 0.0%	0 0.0%	0 0.0%	448 6.2%	4 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.1% 0.9%
	1000IMRF	0 0.0%	0 0.0%	0 0.0%	2 0.0%	446 6.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.6% 0.4%
	1000IRBF	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	2 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.3% 0.7%
	1000NS	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1000ORBF	3 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.3% 0.7%
	1500BBF	9 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	440 6.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.0% 2.0%
	1500CBF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1500CRF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1500ERF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1500IMRF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1500IRBF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	1500NS	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	0 0.0%	0 0.0%	100% 0.0%
	1500ORBF	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	450 6.3%	99.6% 0.4%
	1000BBF	97.1% 2.9%	100% 0.0%	100% 0.0%	99.6% 0.4%	99.1% 0.9%	100% 0.0%	100% 0.0%	100% 0.0%	97.8% 2.2%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.6% 0.4%
		1000BBF	1000CBF	1000CRF	1000ERF	1000IMRF	1000IRBF	1000NS	1000ORBF	1500BBF	1500CBF	1500CRF	1500ERF	1500IMRF	1500IRBF	1500NS	1500ORBF	1000BBF	
		Stvarna klasa																	

Slika 6.16 Matrica zabune evaluacije najboljeg modela

Prema matrici zabune, moguće je utvrditi kako model temeljen na dubokoj konvolucijskoj neuronskoj mreži ispravno klasificira uzorke svih klasa izuzev sljedećih:

- klase 1000BBF koja ispravno predviđa u 98,6 % sa odzivom 97,1 %,
- klase 1500BBF koja ispravno predviđa u 98 % slučajeva sa odzivom 97,8 %,

- klase 1000ERF koja ispravno predviđa u 99,1 % slučajeva sa odzivom 99,6 % i
- klase 1000IMRF koja ispravno predviđa 99,6 % uzoraka sa odzivom klase od 99,1 %.

Dodatno, svojstvo preciznosti klase niže od 100 % izračunano je za klasa 1000IMRF, 1000IRBF te 1500ORBF.

Najniže vrijednosti preciznosti klase (98,6 % i 98 %) i odziva (97,1 % i 97,8 %) izračunate su za klase 1000BBF i 1500BBF, pri čemu je iz matrice zabune vidljivo da je većina pogrešno predviđenih vrijednosti klase 1500BBF zapravo stvarne klase 1000BBF, a isto tako većina pogrešno predviđenih vrijednosti klase 1000BBF je stvarne klase 1500BBF.

Tablica 6.8 sa rezultatima uspješnosti klasifikacije unutar petlje Bayesovog optimizatora, jasno prikazuje kako su sve iteracije sa uspješnošću klasifikacije većom od 99 % rezultat postigle na donjoj granici varijabli $vKoeff_učenja$ (0,005) i $vMomentum$ (0,85). Sukladno tome, s ciljem otkrivanja potencijala prostora rješenja ispod prethodno određenih granica (vidi tablicu 6.4) provedena je dodatna optimizacija sa proširenim rasponima optimizacijskih varijabli $vKoeff_učenja$ i $vMomentum$, prema vrijednostima u tablici 6.9. Ostalim optimizacijskim varijablama rasponi nisu mijenjani.

Tablica 6.9 Prošireni rasponi optimizacijskih varijabli Bayesovog optimizatora

Optimizacijska varijabla	Raspon	Tip podatka
$vKoeff_učenja$ (vKU)	0,001 – 0,01	Kontinuirani
$vMomentum$ (vM)	0,80 – 0,95	Kontinuirani

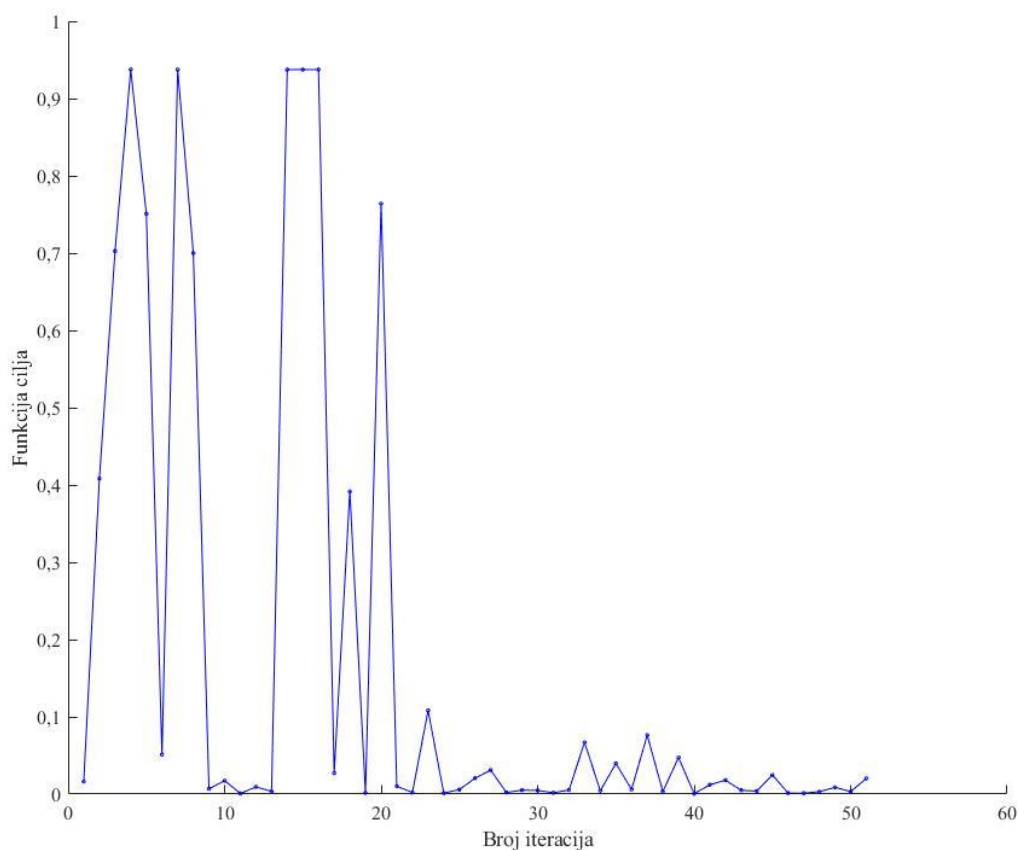
Na temelju proširenih raspona, izvršena je dodatna optimizacija hiperparametara korištenjem hiperparametara Bayesovog optimizatora iz Tablica 6.66.6, uz dodatni uvjet zaustavljanja učenja ukoliko pogreška funkcije cilja nakon 50 izvršenih iteracija bude manja od 0,4 %. Razlog postavljanja uvjeta zaustavljanja na vrijednost 0,4 % jest potencijalno ranije otkrivanje boljeg rješenja od rješenja dobivenih inicijalno provedenom optimizacijom. Tablica 6.10 prikazuje korištene vrijednosti optimizacijskih parametara i izračunatu uspješnost klasifikacije na skupu za testiranje hiperparametara za provedenu dodatnu Bayesovu optimizaciju.

Tablica 6.10 Iteracije dodatne Bayesove optimizacije MVD

Iter. br.	Parametri bloka			Parametri mreže		vFS	UK, [%]
	vDS	vVJ	vBJ	vKU	vM		
1	3	13	15	0,004832198	0,810068554	2	98,37
2	0	10	24	0,008438326	0,805790384	1	59,17
3	4	8	13	0,006617654	0,81432775	2	29,74
4	3	24	23	0,007341955	0,832511218	2	6,25
5	0	43	7	0,007469148	0,808443014	2	24,92
6	7	31	23	0,005369766	0,801017801	2	94,88
7	3	38	6	0,009474643	0,826798278	2	6,25
8	1	30	13	0,005854319	0,880203442	2	30,00
9	4	43	15	0,004763496	0,840992217	3	99,33
10	8	51	12	0,003226666	0,899127572	3	98,25
11	5	41	9	0,002256684	0,867619167	2	99,92
12	6	62	24	0,002071641	0,884170254	2	99,07
13	8	32	17	0,002613947	0,888105595	2	99,64
14	6	46	6	0,009587023	0,804057321	1	6,25
15	3	7	22	0,005377614	0,843773343	3	6,25
16	2	63	22	0,009684139	0,888499679	2	6,25
17	1	10	22	0,002643343	0,809544695	2	97,28
18	5	5	18	0,004650869	0,883883314	2	60,83
19	0	37	20	0,002478562	0,828246491	2	99,84
20	1	53	23	0,00563174	0,881866833	3	23,59
21	4	61	20	0,004252996	0,800301315	2	98,99
22	1	51	6	0,003628579	0,803349127	3	99,78
23	4	7	24	0,003519472	0,801468219	1	89,19
24	2	31	7	0,003962292	0,828440617	2	99,86
25	3	64	21	0,004651064	0,800198499	3	99,40
26	5	10	22	0,005119861	0,800133289	1	97,94
27	1	39	24	0,004827107	0,802363498	2	96,88
28	8	41	6	0,004120916	0,809869142	3	99,78
29	1	43	5	0,002183603	0,850540113	3	99,48
30	2	63	5	0,004391519	0,839675239	1	99,52
31	1	51	6	0,003161369	0,85411704	1	99,82
32	1	15	9	0,002153986	0,800794777	2	99,46
33	2	6	23	0,003524326	0,858393124	3	93,31
34	5	55	10	0,004645494	0,820631799	1	99,56
35	7	64	18	0,003990746	0,844060954	3	96,03
36	3	21	5	0,002838941	0,860680427	3	99,40
37	0	25	4	0,002149194	0,893295458	1	92,38

<i>Iter. br.</i>	<i>Parametri bloka</i>			<i>Parametri mreže</i>			
	<i>vDS</i>	<i>vVJ</i>	<i>vBJ</i>	<i>vKU</i>	<i>vM</i>	<i>vFS</i>	<i>UK, [%]</i>
38	3	62	5	0,002069098	0,832026347	1	99,66
39	0	9	6	0,002017998	0,897774739	3	95,28
40	2	45	4	0,003245109	0,827834971	3	99,94
41	0	63	6	0,002582555	0,898122281	1	98,79
42	6	5	8	0,002026254	0,804961428	3	98,19
43	2	64	5	0,005387638	0,802492044	1	99,50
44	5	28	5	0,004835888	0,805107465	1	99,60
45	3	63	4	0,004041822	0,895404249	1	97,52
46	3	63	5	0,003288788	0,8914224	2	99,86
47	3	61	5	0,002086913	0,81116995	2	99,88
48	1	62	4	0,004765946	0,816069418	3	99,68
49	1	62	4	0,004576231	0,839625301	2	99,13
50	3	59	6	0,002691693	0,871048681	3	99,66
51	0	46	5	0,003752409	0,89991416	2	97,98

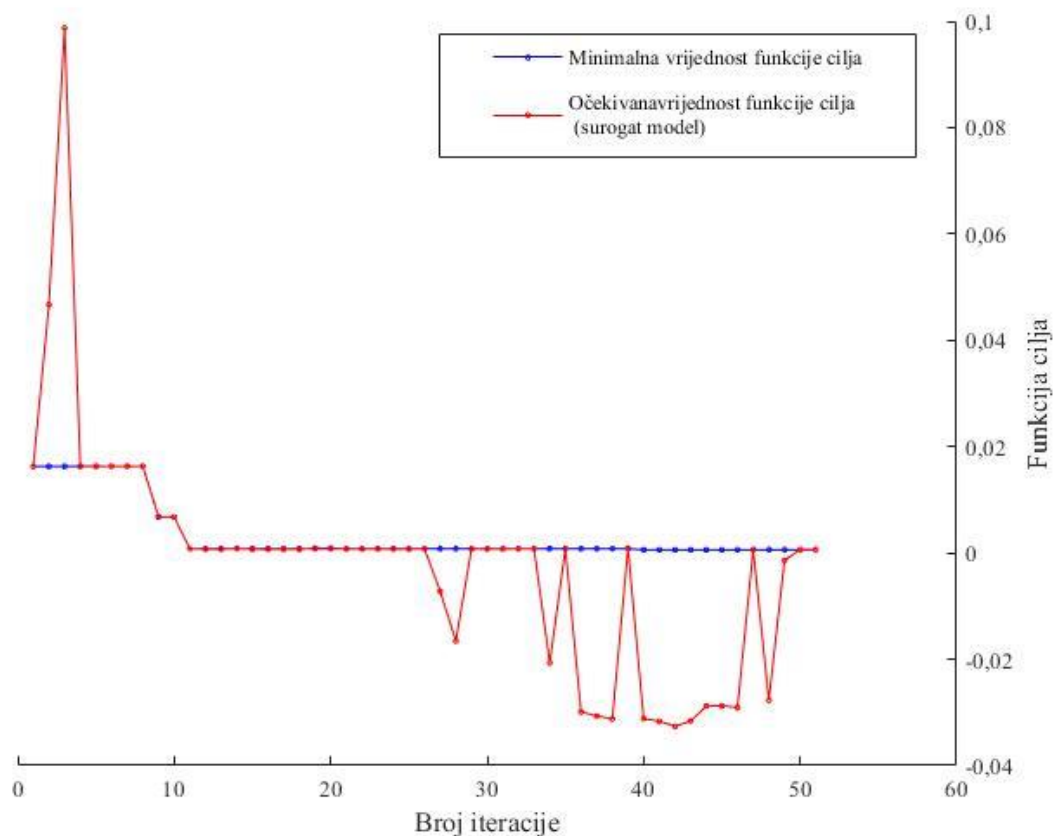
Sukladno prethodno predstavljenim rezultatima, kreiran je grafički prikaz (slika 6.17) izračunate funkcije cilja za svaku evaluaciju hiperparametara (iteraciju) Bayesovog optimizatora.



Slika 6.17 Trend funkcije cilja tijekom dodatne optimizacije

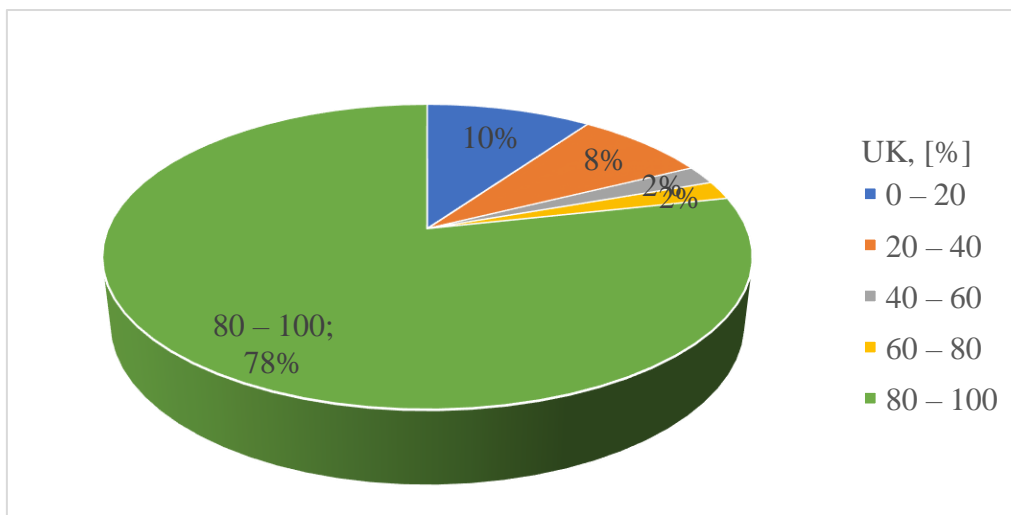
Proširivanje raspona varijabli koje definiraju vrijednosti hiperparametara koeficijenta učenja i momentuma donijelo je povećanje prosječne uspješnosti klasifikacije po iteracijama, što je posljedično generiralo i skup hiperparametara sa najboljom uspješnosti klasifikacije od 99,94 % na skupu za testiranje hiperparametara, što je poboljšanje od 0,34 % u odnosu na prvotno dobiveno rješenje.

Također, kreiran je i grafikon minimalne pogreške funkcije cilja i očekivane pogreške izračunane uz pomoć funkcije stjecanja (vidi sliku 6.18), na kojem je vidljivo kako surogat model Bayesovog optimizatora već u četvrtoj iteraciji uspijeva smanjiti očekivanu vrijednost funkcije surogat modela. Nakon osme iteracije uočavaju se rezultati treniranja Bayesovog optimizatora, pri čemu je vidljiv pad očekivane pogreške i minimalne vrijednosti pogreške funkcije cilja.

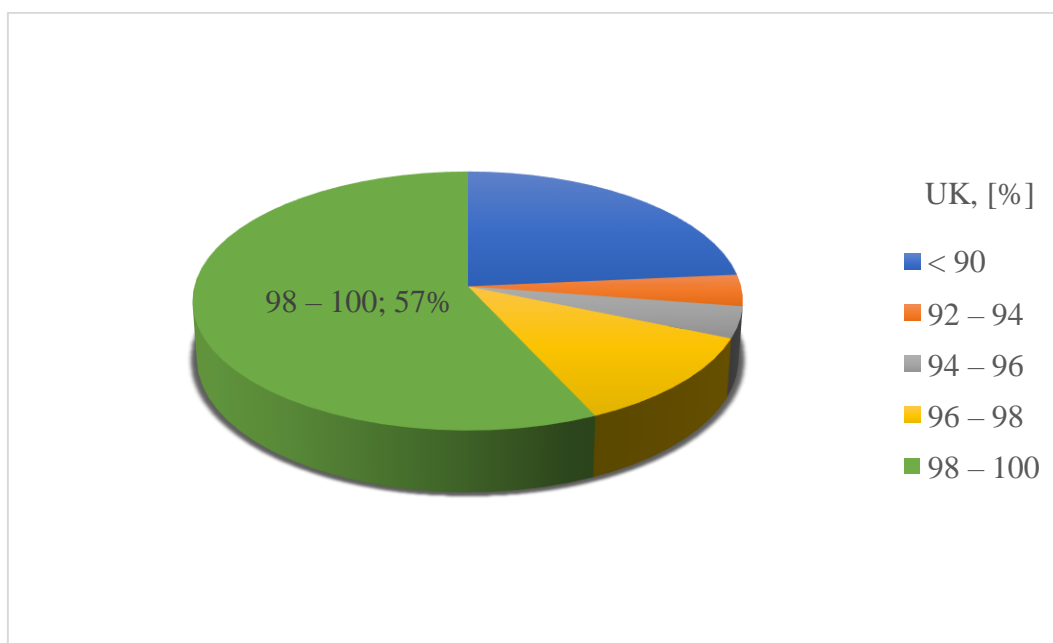


Slika 6.18 Grafikon promjene vrijednosti minimalne i očekivane funkcije cilja tijekom dodatne optimizacije

Nadalje, ponovno je kreiran i grafikon rezultata uspješnosti klasifikacije, pri čemu su rezultati grupirani u intervale širine 20 % (slika 6.19) kao i grafikon rezultata (slika 6.20) u kojem su vrijednosti 90 – 100 % sumirane u grupe od 2 %, a preostale (niže) sumirane vrijednosti uspješnosti klasifikacije pripadaju jednoj grupi (< 90).

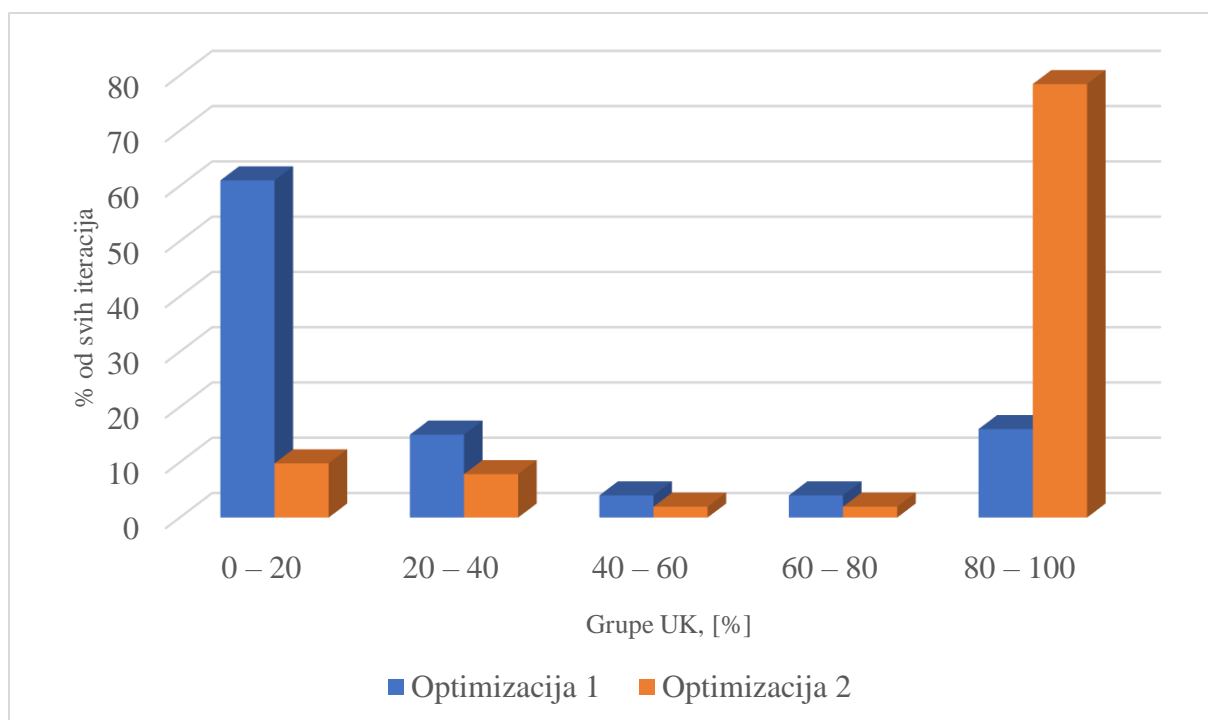


Slika 6.19 Grafikon rezultata UK, [%] - dodatna optimizacija



Slika 6.20 Grafikon grupiranih rezultata UK[%] – dodatna optimizacija

Rezultati prikazani navedenim grafikonima ocrtavaju poboljšanja uspješnosti klasifikacije modela za dodatne iteracije Bayesovog optimizatora uz primjenu povećanih raspona varijabli koeficijenta učenja i momentuma. Grafikonom (slika 6.21) su prikazani uspoređeni rezultati kreiranih grupa rezultata uspješnosti klasifikacije za inicijalnu (Optimizacija 1) i dodatnu optimizaciju (Optimizacija 2), pri čemu je jasno uočljivo kako u dodatnoj optimizaciji dominiraju modeli s više od 80 % uspješnosti klasifikacije, dok u inicijalno provedenoj optimizaciji dominiraju modeli sa uspješnosti klasifikacije nižom od 40 %.



Slika 6.21 Grafikon usporedbe rezultata provedenih optimizacija

Prema dijagramu procesa (slika 6.9), mrežu sa najboljim rezultatom uspješnosti klasifikacije unutar Bayesovog optimizatora provjerava se na testnom skupu podataka. Uspješnost klasifikacije na testnom skupu podataka iznosi 99,94 % odnosno 4 pogrešno klasificirana testna uzorka na ukupno 7 200 testnih mjerenja. Visoki rezultat na skupu za učenje popraćen jednako uspješnim rezultatom klasifikacije testnih primjera dokazuje kako model nema neizraženih svojstava podnaučenosti ili prenaučivosti, ali svakako treba uzeti u obzir i kontekst korištenja laboratorijskog postava (bez uključivanja šumova i realnih uvjeta rada) te testiranje na razmjerno malom udjelu uzoraka.

U nastavku, na slici 6.22, prikazana je matrica zabune za evaluirani model duboke konvolucijske neuronske mreže s najboljim rezultatom na skupu za testiranje hiperparametara (*iteracija br. 40*).

Predviđena klasa	1000BBF	450	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	99.3%
		6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%
	1000CBF	0	450	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
		0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1000CRF	0	0	450	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
		0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1000ERF	0	0	0	450	0	0	0	0	0	0	0	0	0	0	0	0	100%
		0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1000IMRF	0	0	0	0	450	0	0	0	0	0	0	0	0	0	0	0	100%
		0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1000IRBF	0	0	0	0	0	450	0	1	0	0	0	0	0	0	0	0	99.8%
		0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%
	1000NS	0	0	0	0	0	0	450	0	0	0	0	0	0	0	0	0	100%
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1000ORBF	0	0	0	0	0	0	0	450	0	0	0	0	0	0	0	0	100%
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1500BBF	0	0	0	0	0	0	0	0	446	0	0	0	0	0	0	0	100%
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	1500CBF	0	0	0	0	0	0	0	0	0	450	0	0	0	0	0	0	100%
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1500CRF	0	0	0	0	0	0	0	0	0	0	450	0	0	0	0	0	100%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1500ERF	0	0	0	0	0	0	0	0	0	0	0	450	0	0	0	0	100%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	
1500IMRF	0	0	0	0	0	0	0	0	0	0	0	0	450	0	0	0	100%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	0.0%	
1500IRBF	0	0	0	0	0	0	0	0	0	0	0	0	0	450	0	0	100%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	0.0%	
1500NS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	450	0	100%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	0.0%	
1500ORBF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	450	100%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.3%	0.0%	
1000BBF	100%	100%	100%	100%	100%	100%	100%	100%	99.1%	100%	100%	100%	100%	100%	100%	100%	99.9%	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	
	1000BBF	1000CBF	1000CRF	1000ERF	1000IMRF	1000IRBF	1000NS	1000ORBF	1500BBF	1500CBF	1500CRF	1500ERF	1500IMRF	1500IRBF	1500NS	1500ORBF	1000BBF	
	Stvarna klasa																	

Slika 6.22 Matrica zabune najuspješnijeg modela vremenske domene

Konačno, tablica 6.11 definira vrijednosti hiperparametara optimalne strukture modela, a u tablici 6.12 je prikazana konačna struktura optimalnog modela rane procjene kvarova rotacijske opreme temeljenog na dubokoj konvolucijskoj neuronskoj mreži.

Tablica 6.11 Vrijednosti hiperparametara optimalne strukture modela

Razina hiperparametra	Hiperparametar	Korištene optimizacijske varijable	Optimalna vrijednost
Mreža	Koeficijent učenja	v_{KU}	0,003245109
Mreža	Momentum	v_M	0,827834971
KB1	Veličina jezgre	$v_{VJ} \cdot v_{FS}$	135
	Broj jezgri	v_{BJ}	4
KB2	Veličina jezgre	v_{VJ} / v_{FS}	15
	Broj jezgri	$v_{BJ} \cdot v_{FS}$	12
	Broj naslaganih blokova	v_{DS}	2
KB3	Broj jezgri	v_{BJ}	4

Tablica 6.12 Struktura optimalnog modela vremenske domene temeljenog na dubokoj konvolucijskoj neuronskoj mreži

Broj sloja	Tip sloja	Veličina aktivacije	Ukupni broj parametara za učenje
1	Ulazni sloj	12 800 x 1 x 3	0
2	Konvolucijski sloj	12 800 x 1 x 4	1 624
3	Grupna normalizacija	12 800 x 1 x 4	8
4	ReLU	12 800 x 1 x 4	0
5	Sažimanje maksimumom	12 797 x 1 x 4	0
6	Konvolucijski sloj	12 797 x 1 x 12	732
7	Grupna normalizacija	12 797 x 1 x 12	24

<i>Broj sloja</i>	<i>Tip sloja</i>	<i>Veličina aktivacije</i>	<i>Ukupni broj parametara za učenje</i>
8	ReLU	12 797 x 1 x 12	0
9	Konvolucijski sloj	12 797 x 1 x 12	2 172
10	Grupna normalizacija	12 797 x 1 x 12	24
11	ReLU	12 797 x 1 x 12	0
12	Sažimanje maksimumom	12 797 x 1 x 12	0
13	Konvolucijski sloj	12 797 x 1 x 12	588
14	Grupna normalizacija	12 797 x 1 x 12	24
15	ReLU	12 797 x 1 x 12	0
16	Sažimanje uprosječavanjem	12 797 x 1 x 12	0
17	Potpuno povezani sloj	1 x 12 x 16	2 455 888
18	SoftMax	1 x 1 x 16	0
19	Izlazni sloj (klase)	-	0

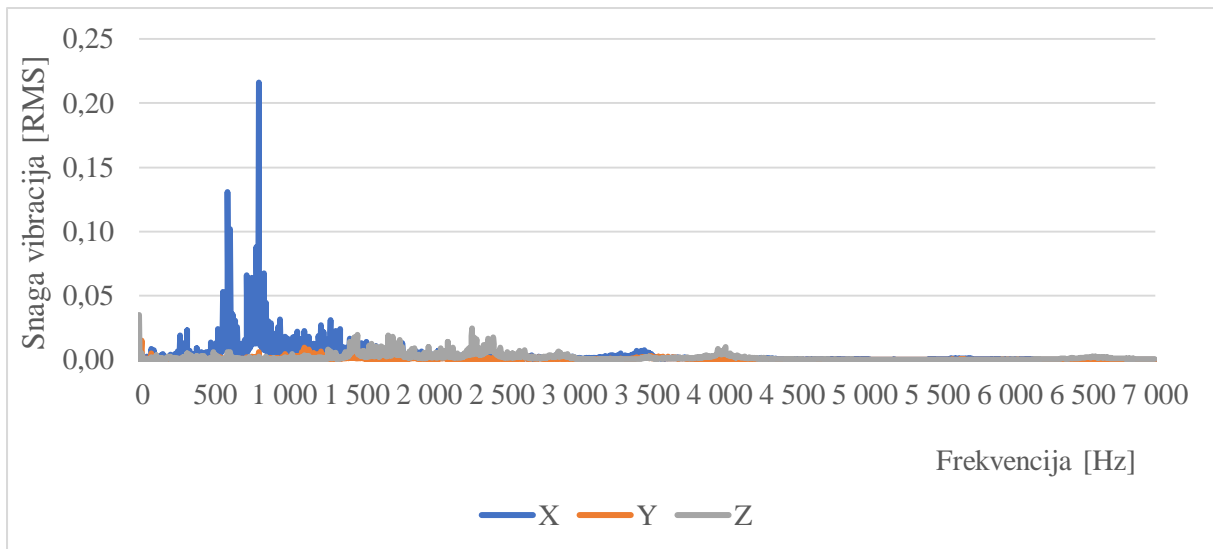
Sukladno tablicama, moguće je utvrditi kako je provedenim postupkom istraživanja opisanim u prethodnom odjeljku, a uz pomoć podataka iz vremenske domene prikupljenih na način opisan u poglavljima (4 i 5) razvijen optimalni model za ranu procjenu uzroka kvarova rotacijske opreme. Razvijeni model vremenske domene u ulazni sloj prihvaća prikupljeni filtrirani uzorak signala sa troosnog senzora akceleracije te u konvolucijskim slojevima automatski izvlači značajke signala za unaprijed definirane klase. Automatski generirane značajke služe kao ulaz u potpuno povezani sloj neuronske mreže, kojim se vrši klasifikacija uzorka. Na temelju provedenih ispitivanja na testnom skupu uzoraka, iznađenom strukturom prikazanom u tablici 6.12 postignuta je uspješnost klasifikacije od 99,94 %.

6.5.2 Model procjene kvarova na temelju signala frekvencijske domene

Učenje modela temeljenog na vremenskoj domeni signala podrazumijeva, kako je opisano u odjeljku 6.2, ulazni signal u obliku *vibracijske slike* dimenzije $1\ 750 \times 1 \times 3$. Svaka *vibracijska slika* pritom sadržava 1 uzorak mjerenja vremenskog signala transformiran FFT algoritmom u frekvencijsku domenu. Postupak snimanja, filtriranja i pretvorbe objašnjen je u odjeljku 5.3.

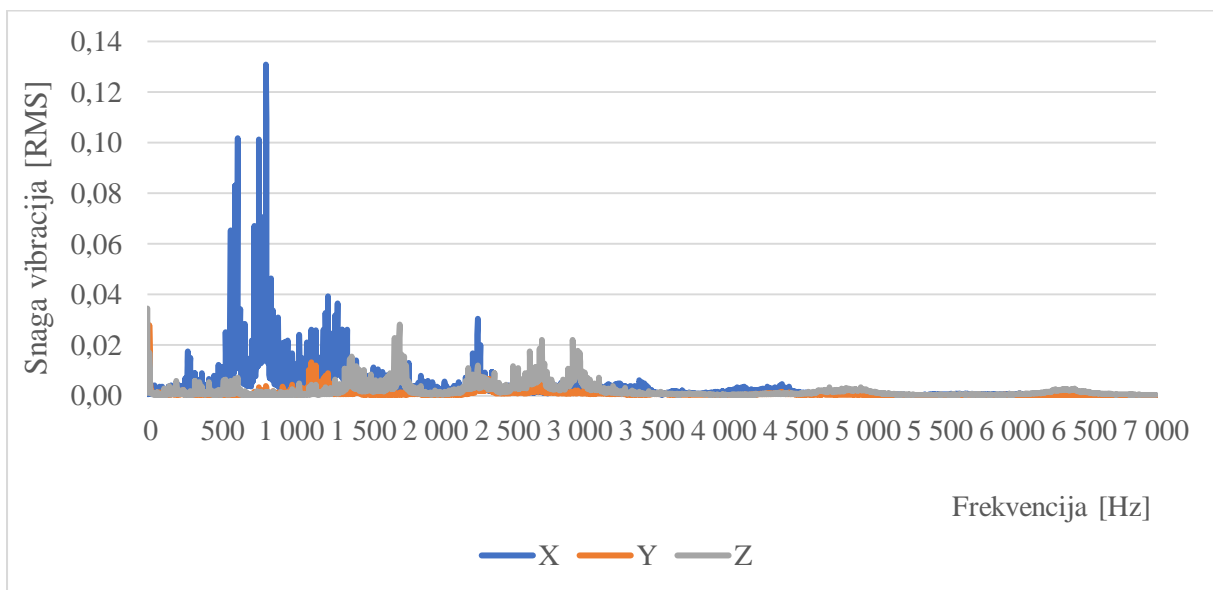
Na temelju pretvorbe FFT algoritmom i filtriranja frekvencija iznad područja rada akcelerometra kreirana je baza uzoraka frekvencijskog područja. Brza Fourierova transformacija kao tehnika pretprocesiranja podrazumijeva postupak izdvajanja značajki. Razvoj optimalnog modela frekvencijske domene temeljenog na dubokoj konvolucijskoj mreži proveden je iz razloga što su inicijalna istraživanja, ali i istraživanja iz literature (tablice 2.1 i 2.2) pokazala kako većina arhitektura standardnih (*plitkih*) umjetnih neuronskih mreža ne doseže uspješnost klasifikacije dubokih struktura bez kvalitetnog izdvajanja vremensko-frekvencijskih značajki. Ovaj model temelji se na korištenju standardnih značajki frekvencijske domene (položaj amplitude na frekvencijskoj skali), koji često predstavlja izlaz sa standardnih IoT vibracijskih senzora sve češće primjenjivanih u okviru koncepta Industrije 4.0. Naime, takvi senzori uslijed ograničene propusnosti infrastrukture na kojoj je izgrađen sustav za prikupljanje podataka, ne omogućavaju u realnom vremenu prijenos signala u vremenskoj domeni, već provode FFT pretvorbu koja definira karakterističan oblik signala koji se šalje u sustav za obradu podataka.

Učenje modela također je izvedeno Bayesovom optimizacijom. Bayesova optimizacija provedena je hiperparametrima zadanim u tablici 6.6 s rasponima vrijednosti varijabli zadanim u tablici 6.4. Primjenjeni su prošireni rasponi varijabli za definiranje hiperparametara koeficijenta učenja i momentuma prema tablici 6.9. Zbog potrebe za kvalitetnijim informiranjem o izgledu frekvencijske karakteristike za pojedine klase, kreirani su grafički prikazi za svaku klasu, od čega su FFT vibracijske slike za dvije klase prikazane u disertaciji. Slika 6.23 prikazuje grafički prikaz karakteristične FFT vibracijske slike za klasu 1000NS. Grafički prikaz izveden je na temelju funkcija za izračun medijana vrijednosti pojedinih frekvencija za X, Y i Z osi za sva mjerenja.



Slika 6.23 Grafički prikaz karakteristične FFT vibracijske slike za klasu 1000NS

Dodatno, kreiran je grafički prikaz za klasu 1000IMRF (slika 6.24) te se na temelju vizualne usporedbe sa grafičkim prikazom frekvencijske domene klase 1000_NS može uvidjeti razlika u vrijednostima i položajima amplituda na pojedinim komponentama (X, Y, Z).



Slika 6.24 Grafički prikaz karakteristične FFT vibracijske slike za klasu 1000IMRF

S obzirom na provedena inicijalna testiranja te provedenu vizualizaciju kao i saznanja o sposobnostima automatskog izdvajanja značajki iz prethodnog odjeljka, postavljen je kriterij ranog zaustavljanja Bayesove optimizacije u slučaju da nakon 30 provedenih optimizacija postoji rješenje sa pogreškom na skupu za testiranje hiperparametara manjom od 0,002. Naime, s obzirom da se FFT procesom već u fazi pretprocesiranja izdvajaju frekvencijske značajke

signala, pretpostavlja se kako će model automatskog izdvajanja značajki uspješno generirati značajke prethodno izračunatih FFT značajki uz manji broj iteracija, nego što je to bio slučaj sa sirovim podacima sa senzora. Tablica 6.13 prikazuje korištene vrijednosti varijabli i vrijednosti uspješnosti klasifikacije provedene optimizacije hiperparametara modela frekvencijske domene. S obzirom na ostvarene vrijednosti, optimizacijski algoritam pokrenuo je uvjet ranog zaustavljanja nakon 30 iteracija. Vidljivo je kako su sve iteracije postigle uspješnost klasifikacije veću od 99 %, a 10 iteracija postiglo je uspješnost klasifikacije na skupu za testiranje hiperparametara od 100 %.

Tablica 6.13 Rezultati optimizacije hiperparametara MFD

Iter. br.	Parametri bloka			Parametri mreže			UK, [%]
	vDS	vVJ	vBJ	vKU	vM	vFS	
2	3	62	13	0,004152137	0,90452518	3	99,90
3	6	29	23	0,004817398	0,86066907	1	99,98
4	6	55	6	0,006690559	0,884084289	1	99,90
5	0	14	17	0,002694797	0,911801522	1	99,15
6	6	55	13	0,006931383	0,80018899	2	99,90
7	5	4	20	0,00721777	0,913262524	2	100,00
8	3	32	23	0,009987604	0,915664306	3	99,92
9	5	59	21	0,00202871	0,928943275	1	99,92
10	8	4	22	0,00567327	0,894392163	2	99,98
11	8	6	7	0,002167327	0,902957645	1	100,00
12	7	5	13	0,002072656	0,900458636	1	100,00
13	3	6	13	0,00699928	0,825308682	3	99,98
14	2	4	8	0,002164676	0,850706283	3	100,00
15	8	4	5	0,00728342	0,887313876	3	99,98
16	0	4	5	0,009880118	0,83285254	3	99,27
17	6	4	10	0,002171139	0,851129424	1	99,98
18	4	4	24	0,008691361	0,886050037	1	99,94
19	8	64	22	0,007269784	0,934990851	1	99,07
20	7	19	19	0,009256109	0,91717873	1	99,92
21	5	32	23	0,009528699	0,899936956	3	100,00
22	2	4	16	0,009165924	0,889914578	3	100,00
23	5	19	18	0,008866336	0,878373951	2	99,72
24	5	44	7	0,005174054	0,879929756	1	99,82
25	8	29	7	0,003212779	0,857180692	3	99,92
26	0	64	23	0,002084002	0,856185802	3	99,15
27	3	4	14	0,006429671	0,861503879	3	100,00

28	6	36	6	0,005269966	0,877807465	2	100,00
29	7	4	13	0,003781165	0,858630514	3	99,98
30	3	19	5	0,009090095	0,840766766	1	100,00

Slijedom dijagrama procesa (vidi sliku 6.9), najbolji model prema metrici uspješnosti klasifikacije testira se na evaluacijskom skupu podataka. U narednom koraku, svih 10 modela frekvencijske domene testirano je na evaluacijskom skupu podataka, a rezultati su dani u tablici 6.14.

Tablica 6.14 Uspješnost klasifikacije MFD na evaluacijskom skupu podataka

Iter. br.	Parametri bloka			Parametri mreže			UK, [%]
	vDS	vVJ	vBJ	vKU	vM	vFS	
7	5	4	20	0,00721777	0,913262524	2	99,94
21	5	32	23	0,009528699	0,899936956	3	99,94
22	2	4	16	0,009165924	0,889914578	3	99,94
12	7	5	13	0,002072656	0,900458636	1	99,97
1	2	19	23	0,009104235	0,912602378	2	99,99
11	8	6	7	0,002167327	0,902957645	1	99,99
14	2	4	8	0,002164676	0,850706283	3	99,99
27	3	4	14	0,006429671	0,861503879	3	99,99
28	6	36	6	0,005269966	0,877807465	2	99,99
30	3	19	5	0,009090095	0,840766766	1	99,99

Iz navedene tablice, moguće je zaključiti kako postoji veći broj kvalitetnih modela frekvencijske domene, odnosno veći broj kombinacija hiperparametara čijim se korištenjem može doći do optimalnog modela, tj. modela koji će ispravno klasificirati uzorke signala prethodno transformiranog u frekvencijsku domenu. Naime, 60 % modela sa uspješnošću klasifikacije od 100 % na skupu za testiranje hiperparametara postiglo je uspješnost klasifikacije od 99,99 % (1 pogrešno klasificiran uzorak) na testnom skupu podataka.

Dosadašnji odjeljci ovog poglavlja opisuju razvojne i evaluacijske stadije modela za ranu procjenu kvarova rotacijske opreme temeljene na konvolucijskim neuronskim mrežama, a u nastavku će biti opisano provedeno testiranje razvijenih i evaluiranih modela. Dodatno testiranje modela provodi se s ciljem otkrivanja robusnosti modela na skupu podataka, koji nije dio skupa podataka korištenih za učenje (strukturiranje) i evaluaciju modela. Dodatno testiranje modela je izvršeno snimanjem i prikupljanjem novog skupa podataka, tj. skupa podataka koji

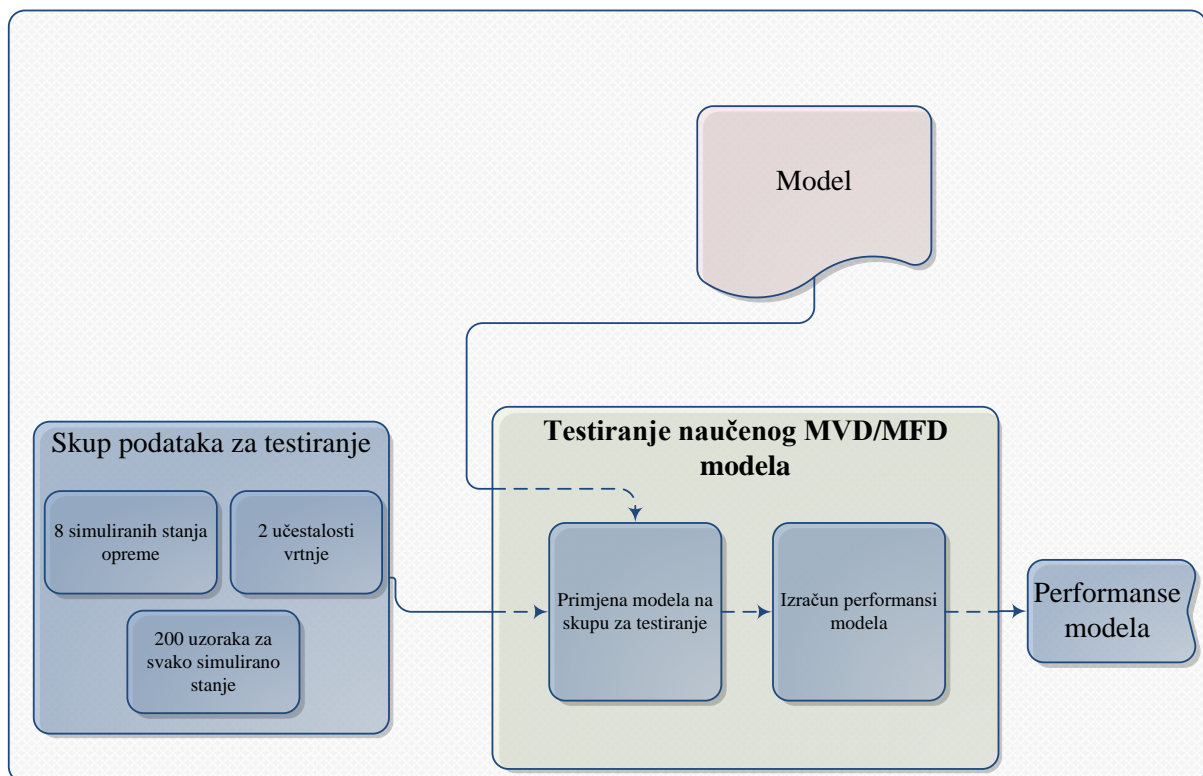
su dosad nepoznati modelu. U zasebno provedenom mjerenju s vremenskim odmakom od mjerenja u kojem su prikupljeni podaci za učenje i evaluaciju, u laboratorijskim uvjetima snimljeno je dodatnih 200 uzoraka za svaki modelom definiran tip kvara. Rezultati testiranja modela vremenske i frekvencijske domene objašnjeni su u narednom odjeljku.

6.5.3 Dodatno testiranje razvijenih modela

Razvijeni i evaluirani modeli vremenske i frekvencijske domene podvrgnuti su dodatnom testiranju, prema procesnoj shemi prikazanoj na dijagramu procesa (slika 6.25).

Testirani su:

- a. model vremenske domene s najboljim performansama na evaluacijskom skupu podataka, opisan u odjeljku 6.5.1, a čiji su hiperparametri definirani u tablici 6.11, testira na testnom skupu podataka, te
- b. modeli frekvencijske domene s najboljim performansama na evaluacijskom skupu, određeni prema tablici 6.14.



Slika 6.25 Proces testiranja naučenog MVD/MFD modela

Naučeni model primjenjuje se na skupu podataka za testiranje na način da se modelu ne definira kojoj klasi pripada uzorak iz skupa podataka, već model na temelju naučenog u prethodnim koracima određuje klasu pojedinog uzorka. Potom se usporedbom stvarnih i predviđenih klasa

određuje uspješnost klasifikacije i definira matrica zabune. Testiranje modela može se procesno usporediti s evaluacijom, no s bitnom razlikom kako se pri testiranju koriste novo prikupljeni podaci. Rezultati testiranja prikazani su u tablici 6.15.

Tablica 6.15 Rezultati testiranja MVD i MFD modela

Testiranje MVD							
	<i>Parametri bloka</i>			<i>Parametri mreže</i>			
<i>Iter. br.</i>	<i>vDS</i>	<i>vVJ</i>	<i>vBJ</i>	<i>vKU</i>	<i>vM</i>	<i>vFS</i>	<i>UK, [%]</i>
40	2	45	4	0,003245109	0,827834971	3	100
Testiranje MFD							
7	5	4	20	0,00721777	0,913262524	2	100
21	5	32	23	0,009528699	0,899936956	3	99,97
22	2	4	16	0,009165924	0,889914578	3	100
12	7	5	13	0,002072656	0,900458636	1	100
1	2	19	23	0,009104235	0,912602378	2	100
11	8	6	7	0,002167327	0,902957645	1	99,88
14	2	4	8	0,002164676	0,850706283	3	100
27	3	4	14	0,006429671	0,861503879	3	100
28	6	36	6	0,005269966	0,877807465	2	99,94
30	3	19	5	0,009090095	0,840766766	1	100

Vidljivo je kako modeli konzistentno zadržavaju vrlo visoke vrijednosti uspješnosti klasifikacije i na testnom skupu podataka, pri čemu MVD postiže maksimalnu vrijednost uspješnosti klasifikacije, dok se vrijednosti UK sa MFD kreću u rasponu 99,88 - 100 %.

Na temelju utvrđenog, može se zaključiti kako razvijeni koncept za iznalaženje hiperparametara modela, kao i pronađeni optimalni skup hiperparametara robusno reagira na nove ulazne podatke, tj. da nema izraženih svojstava prenaučivosti odnosno podnaučivosti.

6.5.4 Vizualizacija razvijenih modela

Modeli konvolucijskih neuronskih mreža smatraju se modelima crne kutije, a sam način izvršavanja modela često je teško shvatljiv. To se posebice odnosi na konvolucijske slojeve, koji su u slučaju ovog istraživanja zaduženi za automatsko učenje značajki ulaznog vibracijskog signala. Tijekom procesa treniranja mreže, kako je objašnjeno u odjeljku 3.1, jezgre

konvolucijskih slojeva uče značajke tj. svojevrsne uzorke labeliranog ulaznog signala u obliku težina.

Nadalje, konvolucijski slojevi razvijenih modela kreiraju matrice aktivacija značajki nad ulaznim signalom čija kombinacija redaka i stupaca predstavlja aktivaciju pojedinih neurona pojedine jezgre na svaki ulazni uzorak signala.

Veličina matrice aktivacija ovisi o tipu sloja konvolucijske mreže:

- broj aktivacija ulaznog sloja jednak je umnošku veličine svakog uzorka ulaznog signala i broja osi akcelerometra,
- broj aktivacija konvolucijskog sloja jednak je umnošku veličine svakog uzorka ulaznog signala i broju jezgri, te
- broj aktivacija potpuno povezanog sloja jednak je zbroju klasa.

Ilustrativno, tablica 6.16 prikazuje broj značajki ulaznog sloja, konvolucijskih slojeva te potpuno povezanog sloja za razvijeni optimalni model vremenske domene. Veliki broj značajki odnosno dimenzija generalno otežava vizualizaciju rezultata rada modela temeljenog na konvolucijskoj umjetnoj neuronskoj mreži.

Tablica 6.16 Broj aktivacija pojedinih slojeva optimalnog MVD

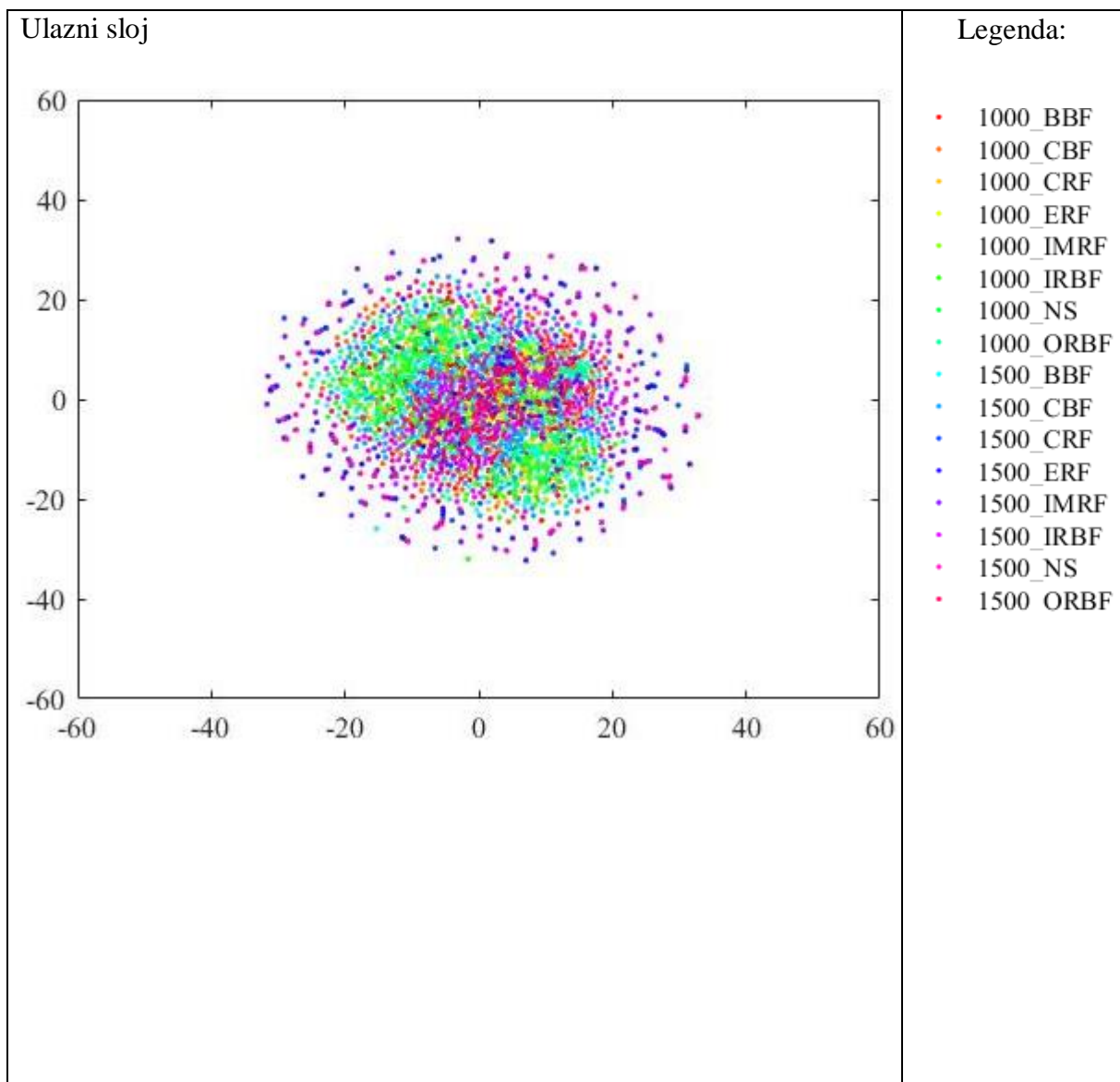
<i>Broj sloja</i>	<i>Tip sloja</i>	<i>Veličina aktivacije</i>	<i>Broj aktivacija</i>
1	Ulazni sloj	12 800 x 1 x 3	38 400
2	Konvolucijski sloj	12 800 x 1 x 4	51 200
6	Konvolucijski sloj	12 797 x 1 x 12	153 564
9	Konvolucijski sloj	12 797 x 1 x 12	153 564
13	Konvolucijski sloj	12 797 x 1 x 12	153 528
17	Potpuno povezani sloj	1 x 12 x 16	192

Manipulacija velikim brojem podataka često već sama po sebi predstavlja izazov. Prirodno, čovjek će se primarno oslanjati na osjetila kako bi barem donekle bio u mogućnosti odgonetnuti pozadinu odnosno bolje shvatiti dostupne mu podatke.

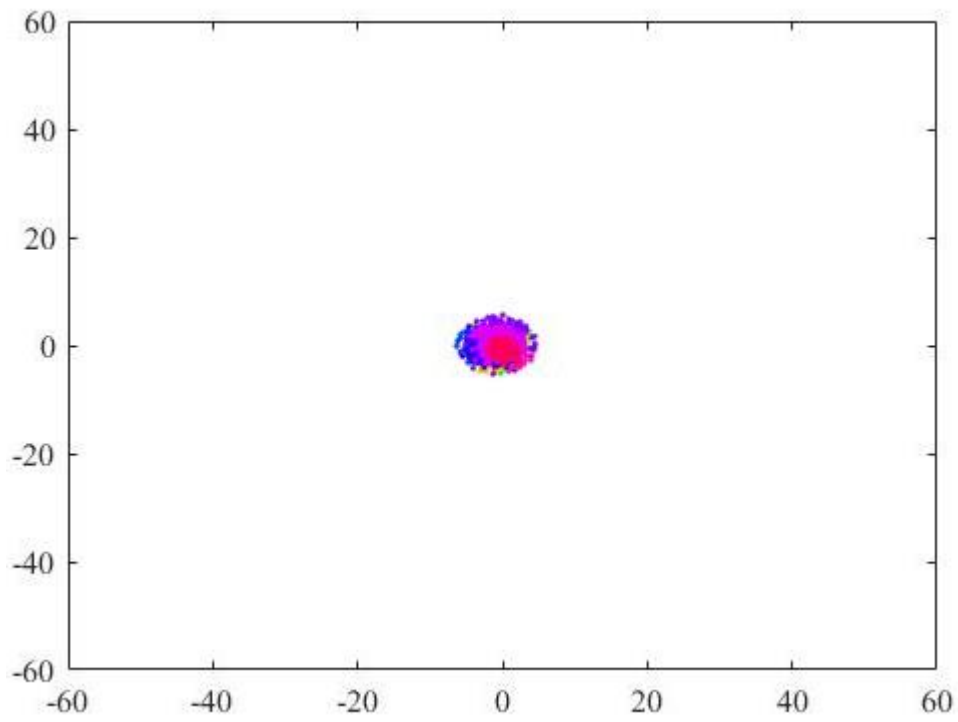
U ovom slučaju, kako bi se aktivacije lakše vizualizirale, primijenit će se metoda za smanjenje broja dimenzija t-SNE (engl. *t-distributed stochastic neighbor embedding*). Ova nelinearna metoda zasniva se na računalno zahtjevnom algoritmu opisanom u [102,103], koji omogućava smanjenje dimenzija kroz izgradnju novog oblika prikaza podataka temeljenog na vjerojatnosnoj ocjeni podobnosti pojedinih uzoraka. Grafički su rezultati metode prikazani za smanjenje broja dimenzija pogodni za dvodimenzionalni prikaz, a na osima su vrijednosti relativnih euklidskih udaljenosti.

Tablica 6.17 vizualizira bezdimenzijske značajke MVD dobivene uz pomoć t-SNE algoritma: reprezentacije značajki za skup podataka za testiranje danih za ulazni sloj (sirovi signal), konvolucijske slojeve i potpuno povezani sloj.

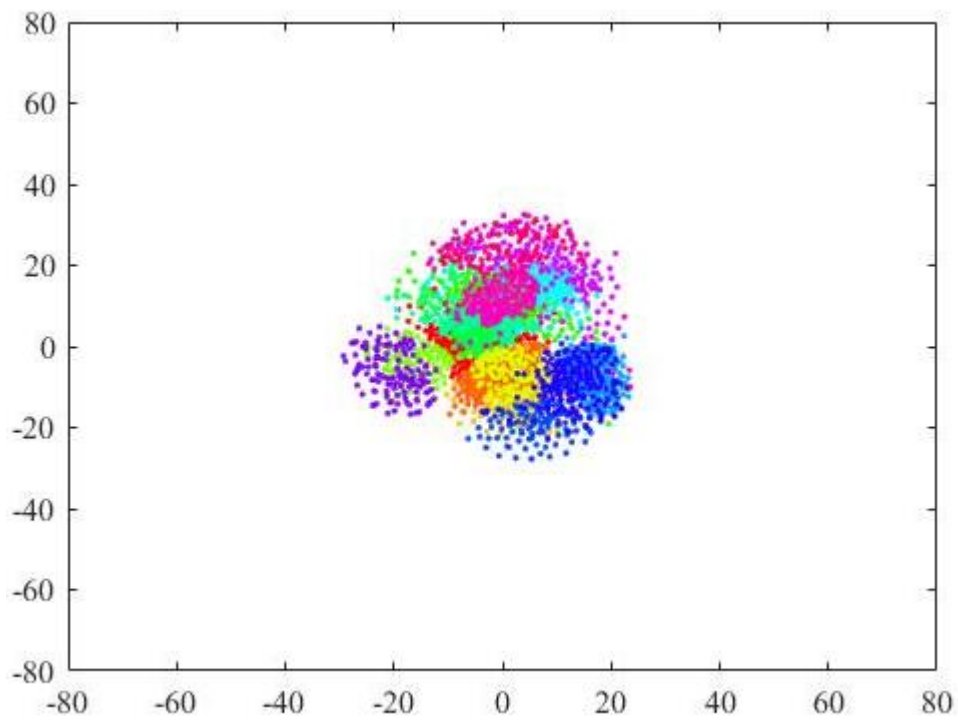
Tablica 6.17 Vizualizacija značajki MVD primjenom t-SNE algoritma



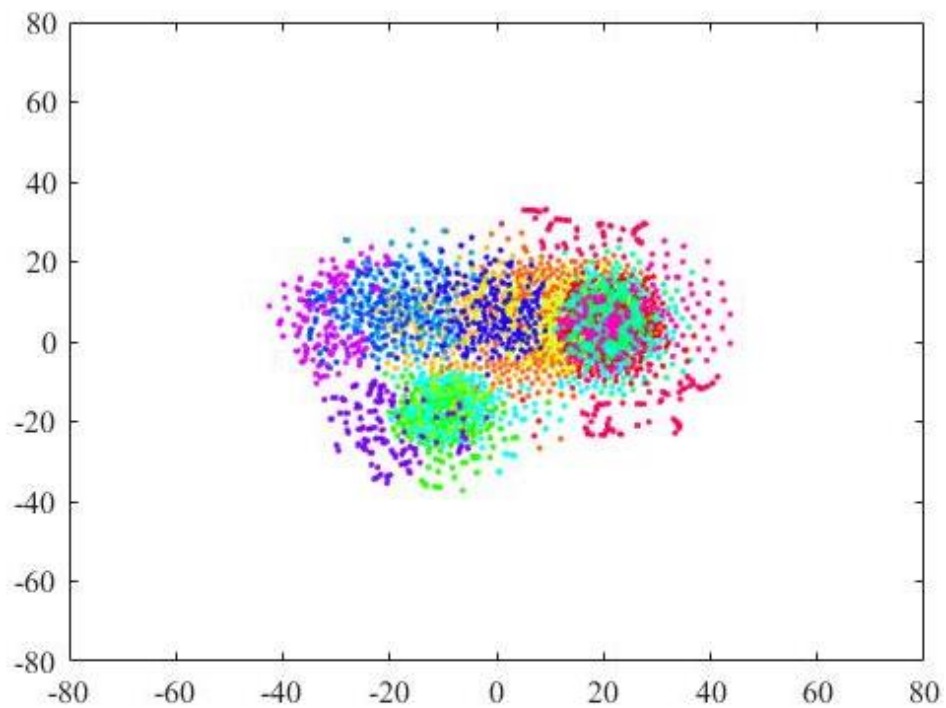
Prvi konvolucijski sloj



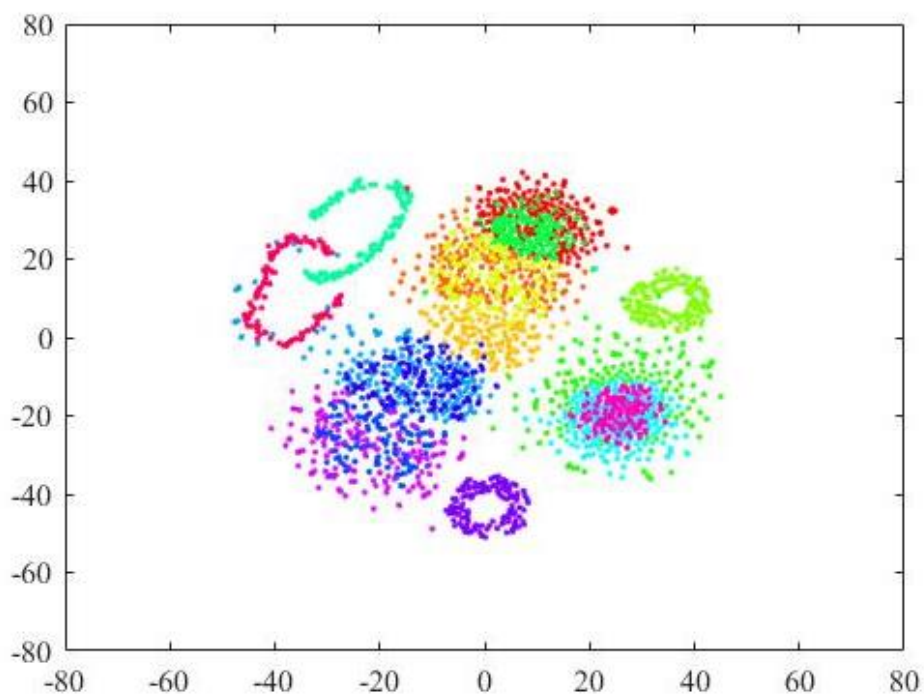
Drugi konvolucijski sloj



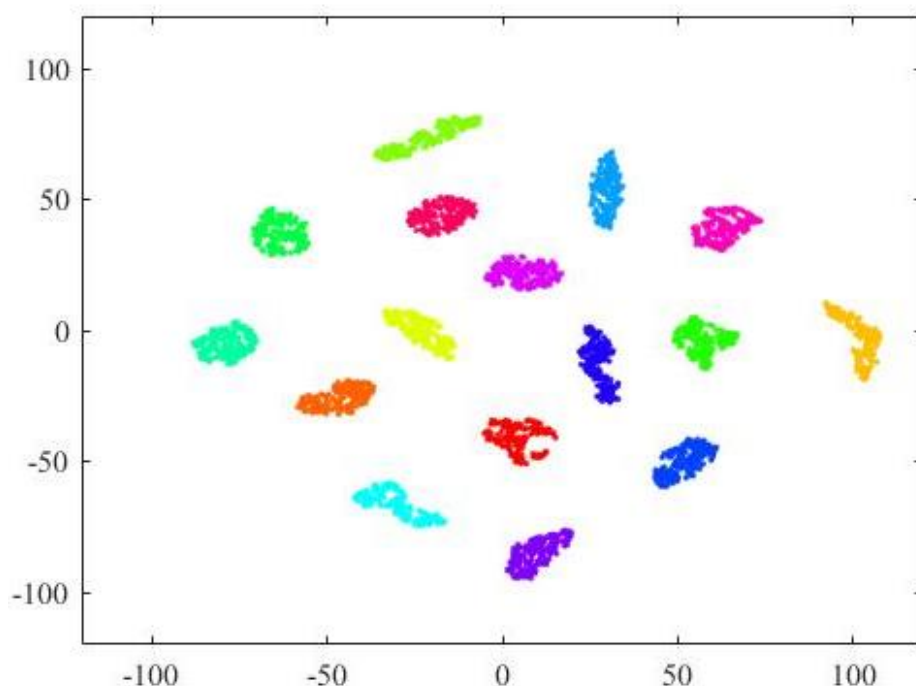
Treći konvolucijski sloj



Četvrti konvolucijski sloj



Potpuno povezani sloj



Na temelju grafičkih prikaza unutar tablice, izvedeni su niže navedeni zaključci.

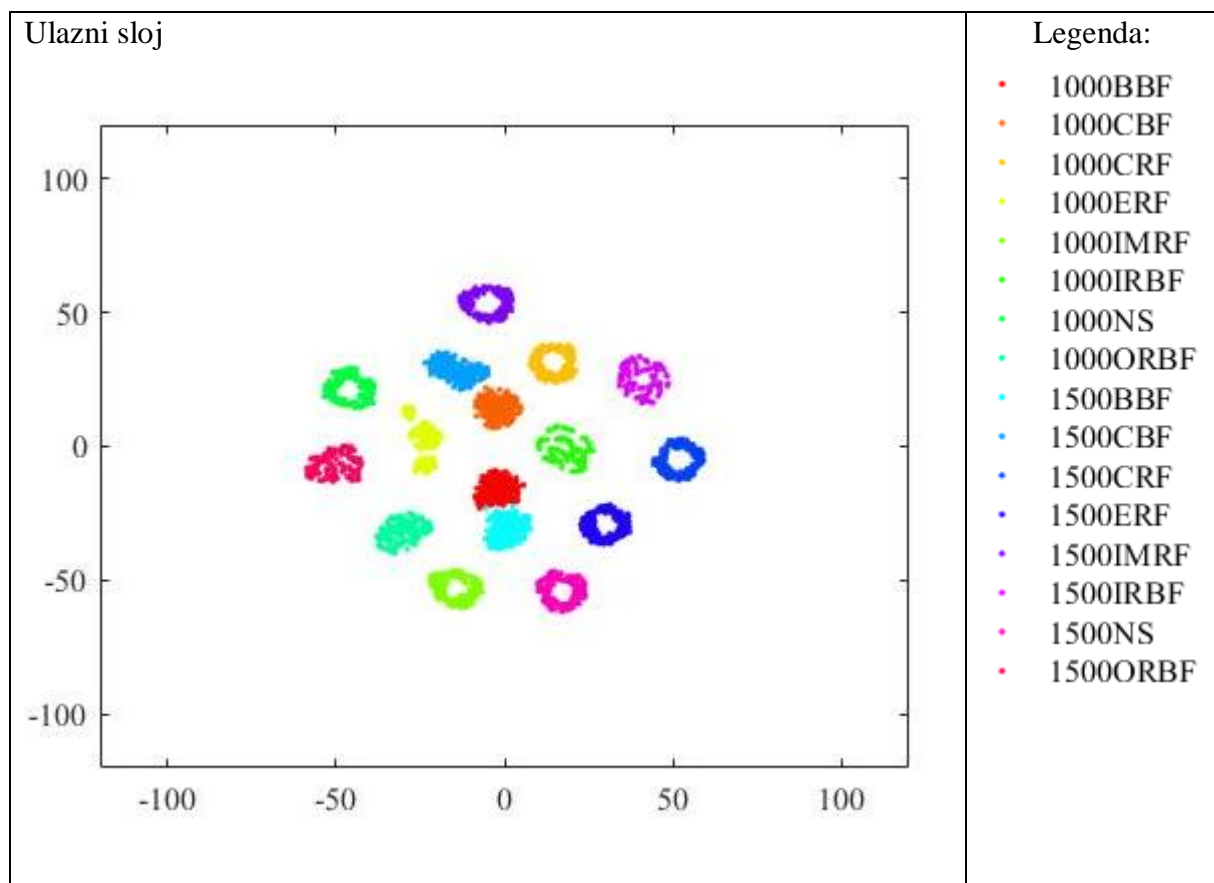
- Značajke koje definiraju pojedine klase jasno se odvajaju kako se prelazi u dublje slojeve modela. Udaljenosti pojedinih točaka u prvom konvolucijskom sloju su male i teško je smisleno grupirati podatke, dok s porastom udaljenosti dolazi do formiranja grupa (klastera) .
- U ranim slojevima podjela po klasama je teško moguća jer dolazi do preklapanja područja pojedinih klasa, no u potpuno povezanom sloju moguća je laka razlikovanje pojedinih klasa na temelju aktivacija značajki.
- Iako postoji jasno razlikovanje između klasa, vizualno nije moguće utvrditi granice pojavnosti između klasa za različite brzine vrtnje.

Osim za MVD, vizualizirane su i značajke MFD za model sa dva konvolucijska sloja (vidi tablicu 6.13: Iter.br 14). Već smanjenjem dimenzija ulaznog sloja moguće je zamijetiti kako je pretvorbom u frekvencijsku domenu brzom Fourierovom transformacijom omogućeno lakše grupiranje signala pojedinih klasa. Ipak, klasteri klasa 1000CBF i 1500CBF te 1000BBF i 1500BBF nalaze se relativno vrlo blizu jedni drugima, a rasponi relativnih udaljenosti

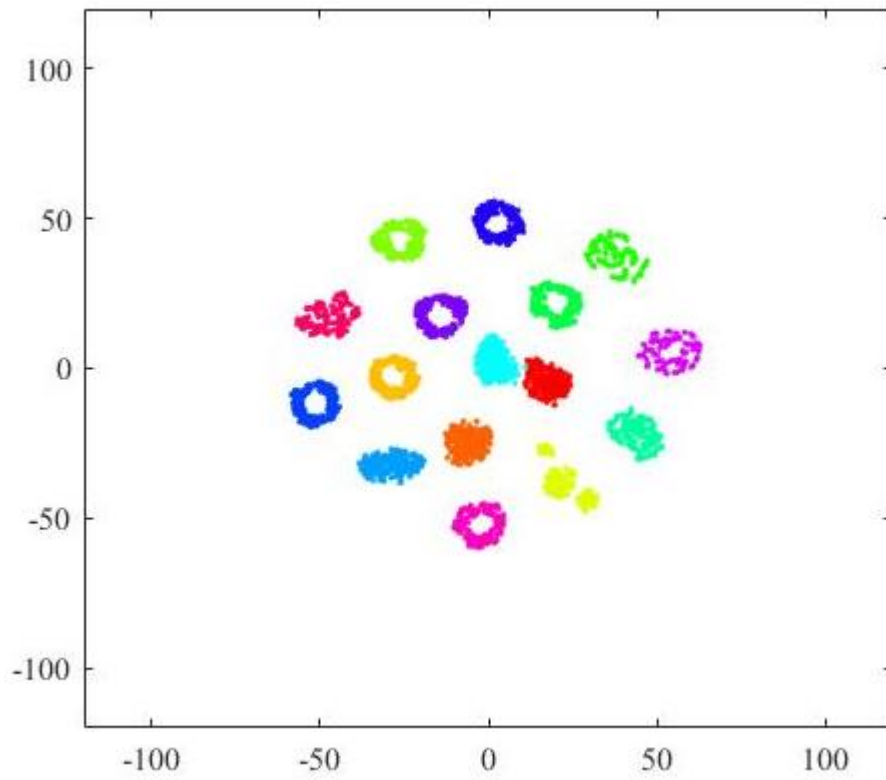
preostalih klasa također su manji. Vizualizacija t-SNE algoritmom (tablica 6.18) vrlo je korisna za ovaj slučaj, jer je vidljivo kako je značajke pojedinih klasa izvučene pretprocesiranjem (FFT pretvorbom) moguće odvojiti nelinearnom reprezentacijom, za razliku od slučaja sa podacima MVD gdje bilo kakva linearna ili nelinearna podjela sirovih podataka po klasama nije jednostavno vidljiva. Prolaskom kroz konvolucijske slojeve, može se zaključiti kako već prvi konvolucijski sloj uspijeva naučiti značajke ulaznog vibracijskog signala. U reprezentaciji potpuno povezanog sloja povećane su relativne udaljenosti između pojedinih grupa klasa, no isto tako vidljivo je vizualno razdvajanje točaka za klasu 1500IRBF.

Na temelju navedenog, postavlja se pitanje o mogućnosti zadane arhitekture neuronske mreže da nauči model isključivo na značajkama frekvencijskog spektra, bez korištenja konvolucijskih slojeva. Za prethodno navedene optimalne hiperparametre (vidi tablicu 6.14), razvijeni su modeli sastavljeni od sljedećih slojeva: ulazni sloj, potpuno povezani sloj, *Softmax* sloj i izlazni sloj, pri čemu je postignut maksimalni rezultat uspješnosti klasifikacije od 92,67 %, kojim se dokazuje opravdanost korištenja konvolucijskih slojeva i u slučaju frekvencijske domene signala kao ulaza u umjetnu neuronsku mrežu.

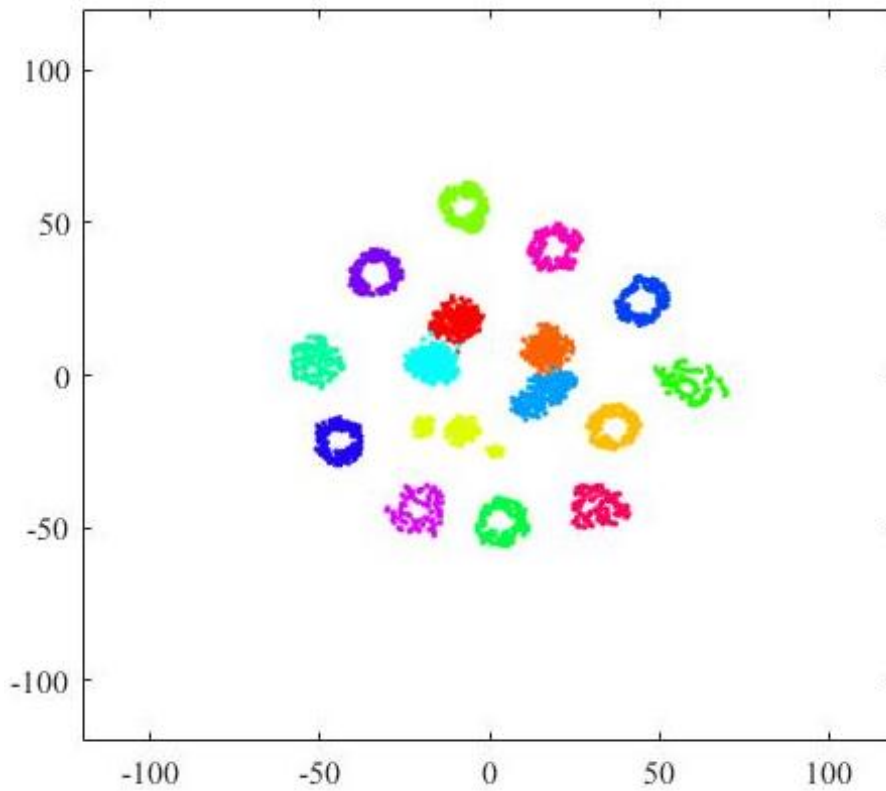
Tablica 6.18 Vizualizacija značajki MFD primjenom t-SNE algoritma



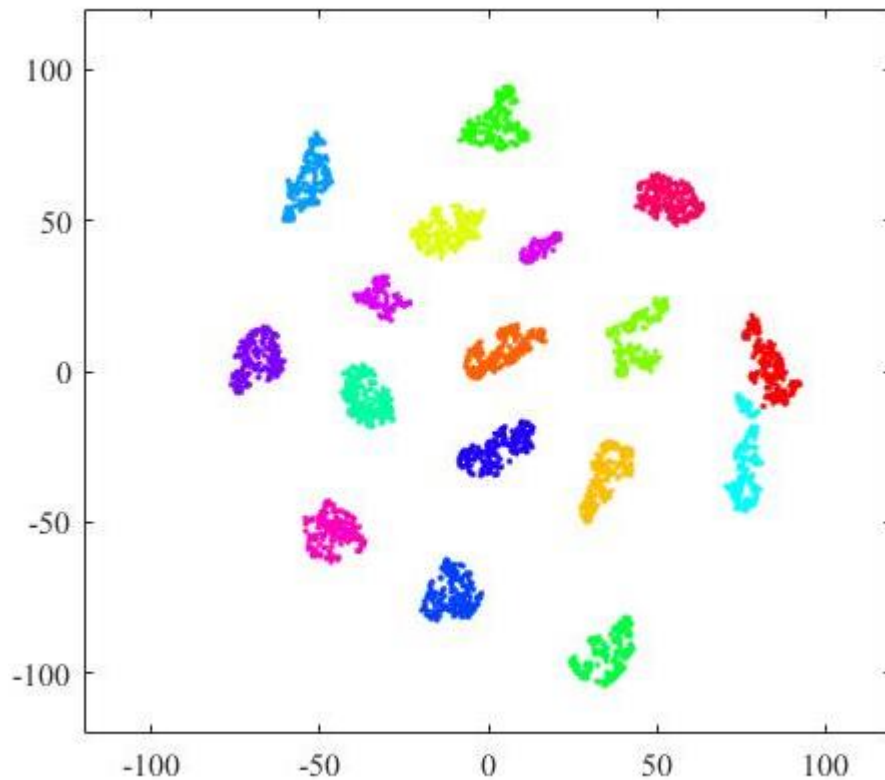
Prvi konvolucijski sloj



Drugi konvolucijski sloj



Potpuno povezani sloj



6.6 Sažetak poglavlja

Na temelju podataka prikupljenih na eksperimentalnom postavu, razvijen je proces sustava za ranu procjenu uzroka kvarova rotacijske opreme. Objašnjena je struktura prikupljenih podataka te način njihove podjele po klasama. Definirana je temeljna struktura dvaju modela temeljenih na vremenskoj, odnosno frekvencijskoj domeni ulaznog signala, pri čemu oba modela koriste arhitekturu dubokih konvolucijskih neuronskih mreža kao osnovu za nadzirano učenje.

Nadalje, pojedinačno su opisani pojedini tipovi slojeva arhitektura te je detaljno objašnjen način učenja modela uz primjenu Bayesove optimizacije hiperparametara. Strukturiran je skup optimizacijskih varijabli, opisan način izračuna hiperparametara konvolucijske neuronske mreže na temelju optimizacijskih varijabli te su određeni rasponi i potrebna ograničenja za primjenu Bayesove optimizacije. Također, navedeni su korišteni hiperparametri algoritma optimizacije.

Prije same provedbe procesa učenja, uspješnost klasifikacije i matrica zabune definirane su kao metrike za evaluaciju modela.

Za proces razvoja modela vremenske domene, izvršena je faza modeliranja sa optimizacijom parametara, pri čemu je izračunano 100 iteracija Bayesove optimizacije s ciljem dobivanja optimalnog modela. Za navedeni proces, izneseni su sljedeći zaključci:

- i. za kreiranje najboljeg modela Bayesovom optimizacijom s uspješnošću klasifikacije od 99,6 % na skupu podataka za testiranje hiperparametara korištene su vrijednosti optimizacijskih varijabli prikazane u tablici 6.19,

Tablica 6.19 Optimizacijske varijable najboljeg MVD prve Bayesove optimizacije

Iter. br.	Parametri bloka KBI			Parametri mreže			
	vDS	vVJ	vBJ	vKU	vM	vFS	UK
82	1	23	5	0,005172443	0,850790306	2	99,60 %

- ii. prema matrici zabune, moguće je utvrditi kako model temeljen na dubokoj konvolucijskoj neuronskoj mreži ispravno klasificira uzorke svih klasa izuzev sljedećih:
 - a. klasa 1000BBF ispravno predviđa u 98,6 % sa odzivom 97,1 %,
 - b. klasa 1500BBF ispravno predviđa u 98 % slučajeva sa odzivom 97,8 %,
 - c. klasa 1000ERF ispravno predviđa u 99,1 % slučajeva sa odzivom 99,6 %,
 - d. klasa 1000IMRF ispravno predviđa 99,6 % uzoraka sa odzivom klase od 99,1 %,
 - e. dodatno, svojstvo preciznosti klase niže od 100 % izračunano je za klase 1000IMRF, 1000IRBF te 1500ORBF,
 - f. najniže vrijednosti preciznosti klase i odziva izračunate su za klase 1500BBF i 1000BBF, pri čemu je iz matrice zabune vidljivo da je većina pogrešno predviđenih vrijednosti klase 1500BBF zapravo stvarne klase 1000BBF, a isto tako većina pogrešno predviđenih vrijednosti klase 1000BBF je stvarne klase 1500BBF, te
- iii. svi rezultati UK viši od 99 % postignuti su na donjim granicama optimizacijskih varijabli $vKoeff_učenja$ i $vMomentum$.

Na temelju iznesenih zaključaka, nastavljen je proces optimizacije za dodatnih 50 iteracija uz proširenje raspona optimizacijskih varijabli $vKoeff_učenja$ i $vMomentum$. Na temelju rezultata dodatne Bayesove optimizacije, izneseni su i dodatni zaključci:

- i. proširivanje raspona varijabli koje definiraju vrijednosti hiperparametara koeficijenta učenja i momentuma donijelo je povećanje prosječne uspješnosti klasifikacije po

iteracijama, što je posljedično generiralo i skup hiperparametara sa najboljom uspjehnosti klasifikacije od 99,94 % na skupu za testiranje hiperparametara, te

- ii. optimalni model vremenske domene postignut je primjenom optimizacijskih varijabli za kreiranje hiperparametara prikazanih u tablici 6.20, te

Tablica 6.20 Optimizacijske varijable najboljeg MVD

Iter. br.	Parametri bloka KBI			Parametri mreže			UK
	vDS	vVJ	vBJ	vKU	vM	vFS	
40	2	45	4	0,003245109	0,827834971	3	99,94%

- iii. provjerom modela na evaluacijskom skupu podataka utvrđena je uspješnost klasifikacije od 99,94 %.

Odjeljak 6.5.2 opisuje razvoj modela frekvencijske domene. Podaci preprocesuirani brzom Fourierovom transformacijom korišteni su za razvoj i evaluaciju modela uz primjenu Bayesove optimizacije hiperparametara konvolucijske neuronske mreže. Na temelju provedenog razvoja, može se zaključiti da:

- i. signal predprocesiran FFT-om već sadrži dio izvučenih značajki te su, prema predviđanjima, brže dosegnute visoke vrijednosti uspjehnosti klasifikacije na skupu za testiranje hiperparametara,
- ii. Postoji veći broj kvalitetnih modela frekvencijske domene, odnosno veći broj hiperparametara čijom se primjenom može doći do optimalnog modela, tj. modela koji će ispravno klasificirati uzorke signala prethodno transformiranog u frekvencijsku domenu, i
- iii. 60 % modela sa uspješnošću klasifikacije od 100 % na skupu za testiranje hiperparametara postiglo je uspješnost klasifikacije od 99,99 % (1 pogrešno klasificiran uzorak) na evaluacijskom skupu podataka.

S ciljem otkrivanja robusnosti modela odnosno otpornosti na svojstva prenaučivosti i podnaučivosti, razvijeni modeli testirani su na skupu podataka snimljenim dodatnim eksperimentom s vremenskim odmakom od početnog skupa. Utvrđeno je kako modeli stabilno zadržavaju vrlo visoke vrijednosti uspjehnosti klasifikacije i na testnom skupu podataka, pri čemu MVD postiže maksimalnu vrijednost uspjehnosti klasifikacije, dok se vrijednosti UK sa MFD kreće u rasponu od 99,88 - 100%.

Dodatno, s ciljem vizualizacije rezultata modela, implementiran je t-SNE algoritam za smanjenje dimenzionalnosti. Na temelju reprezentacije aktivacija konvolucijskih slojeva i potpuno povezanog sloja MVD u nisko-dimenzijskom prostoru na testnim podacima, primijećeno je da:

- i. značajke koje definiraju pojedine klase jasno se odvajaju, kako se prelazi u dublje slojeve modela što znači kako su udaljenosti pojedinih točaka u prvom konvolucijskom sloju male i teško je smisleno grupirati podatke, dok s porastom udaljenosti dolazi do formiranja klastera;
- ii. u ranim slojevima podjela po klasama je teško moguća jer dolazi do preklapanja područja pojedinih klasa, no u potpuno povezanom sloju moguća je lako razlikovanje pojedinih klasa na temelju aktivacija značajki, i
- iii. iako postoji jasna distinkcija između klasa, nije moguće utvrditi granice pojavnosti između klasa za različite brzine vrtnje.

Nadalje, t-SNE algoritam primijenjen je za smanjenje dimenzionalnosti i vizualizaciju aktivacija MFD. Reprezentacija podataka t-SNE modelom ukazala je na mogućnost lakšeg razlikovanja klasa već u ulaznom sloju, što je opravdavalo predviđanja o bržoj konvergenciji modela frekvencijske domene prema optimumu.

Na temelju navedenog, a usko povezano sa rezultatima uspješnosti klasifikacije, moguće je zaključiti kako je moguće koristiti optimizirani MVD za računalno generiranje utjecajnih značajki i procjenu potencijalnih kvarova rotacijske opreme u ranoj fazi njihova nastanka, uz predefinirane klase vibracijskih opterećenja naučene na sirovim podacima troosnog senzora akceleracije.

U slučaju dostupnosti značajki frekvencijske domene generiranih FFT-om, moguće je koristiti optimizirane MFD za računalno učenje značajki već prethodno generiranih značajki, a na temelju kojih je moguće kreirati modele koji također postižu visoke vrijednosti točnosti klasifikacije.

7. ZAKLJUČAK

Ovo poglavlje daje osvrt na definirane istraživačke ciljeve i postavljenu hipotezu istraživanja.

Prikazani su iznađeni zaključci istraživanja i eksperimentalnih rezultata prikazanih u prethodnim poglavljima, te su navedena ograničenja i dane smjernice za nastavak istraživanja.

7.1 Osvrt na znanstvene doprinose i hipotezu istraživanja

Motivacija za izradu ove disertacije proizlazi prvenstveno iz rasta analitičke praznine između mogućnosti prikupljanja podataka u industrijskom okruženju i mogućnosti njihove kvalitetne analize s ciljem potpore ili automatizacije u procesima donošenja odluka. Nit vodilja istraživanja jest ideja o sustavu koji može samostalno procijeniti vlastito stanje odnosno pojavu kvara i pritom koristiti povijesne podatke vlastitog sustava za prikupljanje podataka, kao i podatke sličnih sustava te na osnovu procjene donositi odluke o aktivnostima održavanja. Stoga su u uvodnom dijelu disertacije (odjeljak 1.2) određeni znanstveni doprinosi, koji su i ostvareni ovim istraživanjem:

1. *razvijen model računalnog generiranja utjecajnih značajki vibracijskog signala temeljen na analizi velike količine ulaznih vibracijskih signala i*
2. *razvijen računalni model dubokog strojnog učenja rane procjene kvarova rotacijske opreme temeljen na računalno generiranim utjecajnim značajkama vibracijskih signala.*

Razvijeni modeli za računalno generiranje utjecajnih značajki vibracijskog signala integrirani u računalne modele dubokog strojnog učenja, a s ciljem rane procjene uzroka kvarova rotacijske opreme opisani su u poglavlju 6. Razvijene su strukture modela, koje omogućavaju iskorištavanje podataka sa senzora u sirovom obliku (odjeljak 6.5.1) ili obliku preprocesuiranom brzom Fourierovom transformacijom (odjeljak 6.5.2). Hiperparametri modela su optimizirani Bayesovom optimizacijom, a rezultati provjereni na evaluacijskim skupovima podataka i kasnije testirani na dodatnom skupu podataka. Na kraju, vizualizacijom aktivacija na testnom skupu je utvrđeno uspješno generiranje utjecajnih značajki.

Osim navedenih doprinosa, dodatnim doprinosom može se smatrati prikupljeni skup podataka od više tisuća uzoraka za različite tipove kvarova. Tijekom istraživanja, ukupno je snimljeno više od 100 GB podataka sa troosnog akcelerometra.

Temeljem realiziranih istraživačkih ciljeva i rezultata prezentiranih u poglavlju 6, postavljena hipoteza istraživanja kako je na *temelju ulaznih vibracijskih signala te predefiniranih*

vibracijskih opterećenja, moguće računalno generirati utjecajne značajke te njihovom implementacijom u novorazvijeni računalni model dubokog strojnog učenja, procijeniti potencijalne kvarove rotacijske opreme u ranoj fazi njihova nastanka, je potvrđena.

Hipoteza istraživanja potvrđena je na temelju rezultata razvijenih modela. Vrijednosti definiranih metrika (*uspješnost klasifikacije, matrica zabune*) modela vremenske i modela frekvencijske domene ukazuju kako je moguće na temelju ulaznih vibracijskih signala računalno i bez utjecaja eksperta generirati utjecajne značajke te pomoću njih kreirati model za uspješnu procjenu predefiniраниh tipova kvarova rotacijske opreme.

Razvijeni modeli provjeravani su na skupu podataka snimljenom dodatnim eksperimentom s vremenskim odmakom od početnog skupa. Utvrđeno je kako modeli stabilno zadržavaju vrlo visoke vrijednosti uspješnosti klasifikacije i na testnom skupu podataka, pri čemu model vremenske domene postiže maksimalnu vrijednost uspješnosti klasifikacije od 100 %, dok se vrijednosti uspješnosti klasifikacije sa modelima frekvencijske domene kreće u rasponu od 99,88 - 100 %.

7.2 Ograničenja provedenog istraživanja

Ograničenja istraživanja u ovoj disertaciji povezana su s mogućnostima dostupnog sustava za prikupljanje podataka. Trenutni sustav sastavljen od jednog troosnog vibracijskog senzora u većem industrijskom okruženju može se pokazati neadekvatnim. Iako je prikazani proces razvoja modela skalabilan po broju senzora (vidi tablice 6.2 i 6.3), isti nije testiran na većem broju senzora.

Drugo ograničenje istraživanja vezano je uz mogućnosti korištenog eksperimentalnog postava u simuliranju stupnja kvara. Naime, eksperimentalni postav u ovom obliku uz pomoć nadogradnih modula omogućava simulaciju pojedinih standardnih tipova kvara, ali ne omogućava simuliranje propagacije degradacije stanja kao klasične pojave. S tim povezano, razvijeni modeli omogućavaju uspješno klasificiranje simuliranih standardnih tipova kvarova u ranoj fazi (prije otkaza sustava), ali nisu primjenjivi za praćenje propagacije kvara.

Nadalje, korišteni laboratorijski eksperimentalni postav i provedena mjerenja izoliraju u stvarnom industrijskom okruženju prisutne šumove i ostale poremećaje izazvane radom ostalih strojeva ili prigona razmatranog stroja.

Na kraju, pri implementaciji ovakvih rješenja, uvijek se nameće ograničenje u pogledu kvalitete ulaznih podataka odnosno potrebe osiguravanja kvalitetnog skupa podataka za učenje.

Razvijeni model uspješno će prepoznavati ona stanja koja su mu zadana ulaznim skupom podataka za učenje.

7.3 Smjernice za daljnja istraživanja

Na temelju saznanja otkrivenih u tijeku pisanja ove disertacije te opisanih ograničenja istraživanja, smjernice za buduća istraživanja mogu se podijeliti u nekoliko u nastavku navedenih pravaca:

- *ispitivanje razvijene tehnike na podacima prikupljenima u industrijskom okruženju odnosno podacima sa šumom*

Korišteni laboratorijski eksperimentalni postav i provedena mjerenja izoliraju u stvarnom industrijskom okruženju prisutne šumove i ostale poremećaje izazvane radom ostalih strojeva ili prigona razmatranog stroja te je u daljnjem razvoju potrebno ispitivanje modela odnosno razvijene tehnike u industrijskim uvjetima.

- *prikupljanje novih podataka za različite tipove kvarova i veći raspon broja okretaja za postojeće strukture modela;*

Trenutno razvijeni modeli predefinirani su za procjenu stanja odnosno otkrivanje onih tipova kvarova, za koje su prikupljeni podaci. Simulacija i implementacija dodatnih tipova kvarova doprinijela bi robusnosti modela i osiguranju izlaza za stanja dosad nepoznata modelima.

- *razvoj modela temeljenog na fuziji podataka sa više senzora različitog tipa;*

Iako predviđeni za rad s podacima sa više senzora, trenutni modeli učeni su sa podacima jednog akcelerometra. Proširenje eksperimentalnog postava sa više senzora omogućilo bi provjeru primjenjivosti modela na većem skupu podataka. Također, integracija senzora drugog tipa (npr. senzor akustične emisije) predstavlja jednu od ideja za nastavak istraživanja.

- *razvoj eksperimentalnog postava za mogućnost simuliranja degradacije stanja pojedinih elemenata; te*

Trenutni eksperimentalni postav nema mogućnost simuliranja pojave širenja kvara, koja se činjenično pojavljuje, a također je zamijećena kao nuspojava pri dužim eksperimentima i u ovom istraživanju. Razvoj takvog sustava predstavlja jedan od glavnih željenih smjerova nastavka istraživanja, jer je postojanje takvog sustava imperativ za prilagodbu postojećih modela i implementaciju u realnim sustavima.

- *istraživanje primjene novih tehnika analize podataka temeljenih na dubokim neuronskim mrežama u području održavanja.*

Ovaj smjer istraživanja povezan je sa prethodno navedenim u smislu potrebnih prilagodbi postojećih modela podacima prikupljenim u dužem vremenskom periodu. Naime, trenutni modeli, kao što je dokazano u provedenom istraživanju, pokazuju visoke vrijednosti klasifikacije stanja na temelju automatski generiranih značajki. Pri praćenju stanja u dužem vremenskom periodu, potrebno je u model unijeti i dodatnu varijablu vremena, što podrazumijeva primjenu dodatnih tehnika računalne inteligencije.

U konačnici, a imajući u vidu ciljeve koncepta Industrije 4.0, ostvareni ciljevi provedenog istraživanja kao i ciljevi smjernica za daljnja istraživanja usmjereni su ka razvoju industrijski primjenjivog sustava s mogućnošću samostalne procjene vlastitog stanja, tj. samosvjesnog, samo-održavajućeg sustava koji je sposoban na temelju procjene temeljene na podacima donositi odluke o potrebnim aktivnostima održavanja.

8. LITERATURA

1. N. Maggard B, M. Rhyne D. Total productive maintenance. A timely integration of production and maintenance. *Production and Inventory Management Journal*. 1992.;33(1):6–10.
2. Bevilacqua M, Braglia M. The analytic hierarchy process applied to maintenance strategy selection. *Reliability Engineering & System Safety*. 2000.;70(1):71–83.
3. Mobley RK. An introduction to predictive maintenance. 2nd ed. Amsterdam ; New York: Butterworth-Heinemann; 2002. 438 str.
4. Wireman T. World class maintenance management [Internet]. Industrial Press; 1990. (Creative manufacturing engineering program). Dostupno na: <https://books.google.hr/books?id=JApUAAAAMAAJ>
5. Alsyouf I. Cost effective maintenance for competitive advantages. [Växjö, Sweden]: Växjö University Press; 2004.
6. Chan FTS, Lau HCW, Ip RWL, Chan HK, Kong S. Implementation of total productive maintenance: A case study. *International Journal of Production Economics*. siječanj 2005.;95(1):71–94.
7. Shi J, Wan J, Yan H, Suo H. A survey of cyber-physical systems. U: *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on* [Internet]. IEEE; 2011 [citirano 07. rujan 2017.]. str. 1–6. Dostupno na: <http://ieeexplore.ieee.org/abstract/document/6096958/>
8. Rüssmann M, Lorenz M, Gerbert P, Waldner M, Justus J, Engel P, i ostali. Industry 4.0: The future of productivity and growth in manufacturing industries. Boston Consulting Group [Internet]. 2015. [citirano 08. rujan 2017.];9. Dostupno na: http://www.inovasyon.org/pdf/bcg.perspectives_Industry.4.0_2015.pdf
9. Pfohl H-C, Yahsi B, Kurnaz T. The Impact of Industry Supply Chain. U: *Proceedings of the Hamburg International Conference of Logistics (HICL)* [Internet]. Hamburg; 2015 [citirano 26. rujan 2017.]. Dostupno na: <https://hicl.org/publications/2015/20/31.pdf>
10. Gorecky D, Schmitt M, Loskyll M, Zuhlke D. Human-machine-interaction in the industry 4.0 era. U: *2014 12th IEEE International Conference on Industrial Informatics (INDIN)* [Internet]. Porto Alegre RS, Brazil: IEEE; 2014 [citirano 08. ožujak 2019.]. str. 289–94. Dostupno na: <http://ieeexplore.ieee.org/document/6945523/>
11. ISO 20816-1:2016 - Mechanical vibration -- Measurement and evaluation of machine vibration [Internet]. [citirano 28. rujan 2017.]. Dostupno na: <https://www.iso.org/standard/63180.html>
12. EN 13306. Maintenance Terminology, European Standard. Brussels: CEN (European Committee for Standardization); 2010.
13. Holmberg K, urednik. E-maintenance. London ; New York: Springer; 2010. 511 str.

14. Niu G, Yang B-S, Pecht M. Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. *Reliability Engineering & System Safety*. srpanj 2010.;95(7):786–96.
15. Yu R. *Theory of Predictive Maintenance and its Application within Framework of Intelligent Control-Maintenance-Management System*. Huazhong University of Science and Technology, Wuhan, China. 2000.;
16. Sullivan G, Pugh R, Melendez AP, Hunt WD. *Operations & Maintenance Best Practices - A Guide to Achieving Operational Efficiency (Release 3)* [Internet]. 2010 kol [citirano 28. studeni 2018.]. Report No.: PNNL-19634, 1034595. Dostupno na: <http://www.osti.gov/servlets/purl/1034595/>
17. Bloch HP, Geitner FK. *Machinery failure analysis and troubleshooting*. Fourth edition. Oxford: BH/Elsevier; 2012. 743 str. (Practical machinery management for process plants).
18. Adamović Ž, Jeftić M. *Preventivno održavanje u mašinstvu*. Beograd: Građevinska knjiga; 1988.
19. Wang KS. Key Techniques in Intelligent Predictive Maintenance (IPdM) – A Framework of Intelligent Faults Diagnosis and Prognosis System (IFDaPS). *Advanced Materials Research*. listopad 2014.;1039:490–505.
20. Bengtsson M. *Decision-making During Condition Based Maintenance Implementation*. U Coimbra, Portugal; 2007.
21. Ahmad R, Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering*. kolovoz 2012.;63(1):135–49.
22. Girdhar P, Scheffer C. *Practical machinery vibration analysis and predictive maintenance*. Amsterdam ; Boston : Burlington, MA: Elsevier ; Newnes; 2004. 255 str. (Practical professional books from Elsevier).
23. Liu Y, Guo L, Wang Q, An G, Guo M, Lian H. Application to induction motor faults diagnosis of the amplitude recovery method combined with FFT. *Mechanical Systems and Signal Processing*. studeni 2010.;24(8):2961–71.
24. Wu J-D, Liu C-H. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. *Expert Systems with Applications*. travanj 2009.;36(3):4278–86.
25. Wu J-D, Chen J-C. Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines. *NDT & E International*. lipanj 2006.;39(4):304–11.
26. Butler DE. The Shock-pulse method for the detection of damaged rolling bearings. *Non-Destructive Testing*. travanj 1973.;6(2):92–5.
27. Yoshioka T, Fujiwara T. A new acoustic emission source locating system for the study of rolling contact fatigue. *Wear*. rujan 1982.;81(1):183–6.

28. Al-Dossary S, Raja Hamzah RI, Mba D. Acoustic Emission Waveform Changes for Varying Seeded Defect Sizes. *Advanced Materials Research*. veljača 2006.;13–14:427–32.
29. Gowid S, Dixon R, Ghani S. A novel robust automated FFT-based segmentation and features selection algorithm for acoustic emission condition based monitoring systems. *Applied Acoustics*. veljača 2015.;88:66–74.
30. Al-Obaidi SMA, Leong MS, Hamzah RIR, Abdelrhman AM. A Review of Acoustic Emission Technique for Machinery Condition Monitoring: Defects Detection & Diagnostic. *Applied Mechanics and Materials*. studeni 2012.;229–231:1476–80.
31. Lu B, Durocher D, Stemper P. Predictive maintenance techniques. *IEEE Industry Applications Magazine*. studeni 2009.;15(6):52–60.
32. Zarei J, Tajeddini MA, Karimi HR. Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics*. ožujak 2014.;24(2):151–7.
33. Tahir et al. Extracting accurate time domain features from vibration signals for reliable classification of bearing faults. *International Journal of ADVANCED AND APPLIED SCIENCES*. siječanj 2018.;5(1):156–63.
34. Žvokelj M, Zupan S, Prebil I. Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemble Empirical Mode Decomposition method combined with Principal Component Analysis. *Mechanical Systems and Signal Processing*. svibanj 2010.;24(4):1049–67.
35. Hernández-Muriel JA, Álvarez-Meza AM, Echeverry-Correa JD, Orozco-Gutierrez ÁÁ, Álvarez-López MA. Feature relevance estimation for vibration-based condition monitoring of an internal combustion engine. *TecnoLógicas*. 02. svibanj 2017.;20(39):157–72.
36. Li L, Song Y. An On-line Feature Extraction Method for Transformer Vibration Signals. *DEStech Transactions on Computer Science and Engineering [Internet]*. 27. ožujak 2018. [citirano 04. ožujak 2019.];(cmee). Dostupno na: <http://dpi-proceedings.com/index.php/dtcse/article/view/20031>
37. Mehta P, Werner A, Mears L. Condition based maintenance-systems integration and intelligence using Bayesian classification and sensor fusion. *Journal of Intelligent Manufacturing*. travanj 2015.;26(2):331–46.
38. Al-Badour F, Sunar M, Cheded L. Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques. *Mechanical Systems and Signal Processing*. kolovoz 2011.;25(6):2083–101.
39. Antoniadou I, Manson G, Staszewski WJ, Barszcz T, Worden K. A time–frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions. *Mechanical Systems and Signal Processing*. prosinac 2015.;64–65:188–216.
40. Yan J. *Machinery prognostics and prognosis oriented maintenance management*. Singapore: John Wiley & Sons Singapore Pte. Ltd; 2015. 337 str.

41. Koprinkova-Hristova P, Hadjiski M, Doukovska L, Beloreshki S. Recurrent Neural Networks for Predictive Maintenance of Mill Fan Systems. *International Journal of Electronics and Telecommunications* [Internet]. 01. siječanj 2011. [citirano 06. ožujak 2019.];57(3). Dostupno na: <http://content.sciendo.com/view/journals/eletel/57/3/article-p401.xml>
42. Samanta B, Al-Balushi KR. ARTIFICIAL NEURAL NETWORK BASED FAULT DIAGNOSTICS OF ROLLING ELEMENT BEARINGS USING TIME-DOMAIN FEATURES. *Mechanical Systems and Signal Processing*. ožujak 2003.;17(2):317–28.
43. Rohani A, Abbaspour-Fard MH, Abdolahpour S. Prediction of tractor repair and maintenance costs using Artificial Neural Network. *Expert Systems with Applications*. srpanj 2011.;38(7):8999–9007.
44. Jiu-ling Zhao, Jiu-fen Zhao. Application of BPNN and CBR on Fault Diagnosis for Missile Electronic Command System. U: 2006 6th World Congress on Intelligent Control and Automation [Internet]. Dalian, China: IEEE; 2006 [citirano 06. ožujak 2019.]. str. 5796–9. Dostupno na: <http://ieeexplore.ieee.org/document/1714187/>
45. Samanta B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing*. svibanj 2004.;18(3):625–44.
46. Hwang Y-R, Jen K-K, Shen Y-T. Application of cepstrum and neural network to bearing fault detection. *Journal of Mechanical Science and Technology*. listopad 2009.;23(10):2730–7.
47. Mahanty RN, Dutta Gupta PB. Application of RBF neural network to fault classification and location in transmission lines. *IEE Proceedings - Generation, Transmission and Distribution*. 2004.;151(2):201.
48. Samantaray SR, Dash PK, Panda G. Fault classification and location using HS-transform and radial basis function neural network. *Electric Power Systems Research*. lipanj 2006.;76(9–10):897–905.
49. Xiong G, Shi D, Chen J, Zhu L, Duan X. Divisional fault diagnosis of large-scale power systems based on radial basis function neural network and fuzzy integral. *Electric Power Systems Research*. prosinac 2013.;105:9–19.
50. Al-Raheem KF, Roy A, Ramachandran KP, Harrison DK, Grainger S. Application of the Laplace-wavelet combined with ANN for rolling bearing fault diagnosis. *Journal of Vibration and Acoustics*. 2008.;130(5):051007.
51. Cottrell M, Gaubert P, Eloy C, François D, Hallaux G, Lacaille J, i ostali. Fault Prediction in Aircraft Engines Using Self-Organizing Maps. U: Príncipe JC, Miikkulainen R, urednici. *Advances in Self-Organizing Maps* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009 [citirano 06. ožujak 2019.]. str. 37–44. Dostupno na: http://link.springer.com/10.1007/978-3-642-02397-2_5
52. Miljkovic D. Fault detection for aircraft piston engine using self-organizing map. U: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) [Internet]. Opatija, Croatia: IEEE; 2017

- [citirano 06. ožujak 2019.]. str. 1067–72. Dostupno na: <http://ieeexplore.ieee.org/document/7973582/>
53. von Birgelen A, Buratti D, Mager J, Niggemann O. Self-Organizing Maps for Anomaly Localization and Predictive Maintenance in Cyber-Physical Production Systems. *Procedia CIRP*. 2018.;72:480–5.
 54. Pan Y, Er MJ, Li X, Yu H, Gouriveau R. Machine health condition prediction via online dynamic fuzzy neural networks. *Engineering Applications of Artificial Intelligence*. listopad 2014.;35:105–13.
 55. Junbo T, Weining L, Juneng A, Xueqian W. Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder. U: The 27th Chinese Control and Decision Conference (2015 CCDC) [Internet]. Qingdao, China: IEEE; 2015 [citirano 06. ožujak 2019.]. str. 4608–13. Dostupno na: <http://ieeexplore.ieee.org/document/7162738/>
 56. Zhu Huijie, Rui Ting, Wang Xinqing, Zhou You, Fang Husheng. Fault diagnosis of hydraulic pump based on stacked autoencoders. U: 2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI) [Internet]. Qingdao, China: IEEE; 2015 [citirano 06. ožujak 2019.]. str. 58–62. Dostupno na: <http://ieeexplore.ieee.org/document/7494195/>
 57. Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement*. srpanj 2016.;89:171–8.
 58. Shao H, Jiang H, Zhang X, Niu M. Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology*. 01. studeni 2015.;26(11):115002.
 59. Shao S, Sun W, Wang P, Gao RX, Yan R. Learning features from vibration signals for induction motor fault diagnosis. U IEEE; 2016 [citirano 22. svibanj 2018.]. str. 71–6. Dostupno na: <http://ieeexplore.ieee.org/document/7790138/>
 60. Li Z, Wang Y, Wang K. A data-driven method based on deep belief networks for backlash error prediction in machining centers. *Journal of Intelligent Manufacturing* [Internet]. 19. prosinac 2017. [citirano 06. ožujak 2019.]; Dostupno na: <http://link.springer.com/10.1007/s10845-017-1380-9>
 61. Wang J, Wang K, Wang Y, Huang Z, Xue R. Deep Boltzmann machine based condition prediction for smart manufacturing. *Journal of Ambient Intelligence and Humanized Computing*. ožujak 2019.;10(3):851–61.
 62. Shaheryar A, Yin X-C, Yousuf W. Robust Feature Extraction on Vibration Data under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines. *International Journal of Computer Applications*. 15. lipanj 2017.;167(4):37–45.
 63. Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufer M, Verstockt S, i ostali. Convolutional Neural Network Based Fault Detection for Rotating Machinery. *Journal of Sound and Vibration*. rujan 2016.;377:331–45.

64. Zhang W, Peng G, Li C, Chen Y, Zhang Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors*. 22. veljača 2017.;17(2):425.
65. Li S, Liu G, Tang X, Lu J, Hu J. An Ensemble Deep Convolutional Neural Network Model with Improved D-S Evidence Fusion for Bearing Fault Diagnosis. *Sensors*. 28. srpanj 2017.;17(8):1729.
66. Zhang W, Li C, Peng G, Chen Y, Zhang Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*. veljača 2018.;100:439–53.
67. Hoang D-T, Kang H-J. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research* [Internet]. ožujak 2018. [citirano 06. lipanj 2018.]; Dostupno na: <http://linkinghub.elsevier.com/retrieve/pii/S1389041717303261>
68. Wen L, Li X, Gao L, Zhang Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Transactions on Industrial Electronics*. srpanj 2018.;65(7):5990–8.
69. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016.
70. Hochreiter S. *Untersuchungen zu dynamischen neuronalen Netzen*. Technische Universität München; 1991.
71. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks*. siječanj 2015.;61:85–117.
72. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 24. svibanj 2017.;60(6):84–90.
73. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv:151106434 [cs] [Internet]. 19. studeni 2015. [citirano 07. ožujak 2019.]; Dostupno na: <http://arxiv.org/abs/1511.06434>
74. Dahl GE, Dong Yu, Li Deng, Acero A. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*. siječanj 2012.;20(1):30–42.
75. Hinton G, Deng L, Yu D, Dahl G, Mohamed A, Jaitly N, i ostali. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*. studeni 2012.;29(6):82–97.
76. Noda K, Yamaguchi Y, Nakadai K, Okuno HG, Ogata T. Audio-visual speech recognition using deep learning. *Applied Intelligence*. lipanj 2015.;42(4):722–37.
77. Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. U: Salinesi C, Norrie MC, Pastor Ó, urednici. *Advanced Information Systems Engineering* [Internet]. Berlin, Heidelberg: Springer

- Berlin Heidelberg; 2013 [citirano 07. ožujak 2019.]. str. 403–10. Dostupno na: http://link.springer.com/10.1007/978-3-642-40763-5_50
78. Xu Y, Mo T, Feng Q, Zhong P, Lai M, Chang EI-C. Deep learning of feature representation with multiple instance learning for medical image analysis. U: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [Internet]. Florence, Italy: IEEE; 2014 [citirano 07. ožujak 2019.]. str. 1626–30. Dostupno na: <http://ieeexplore.ieee.org/document/6853873/>
 79. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, i ostali. A survey on deep learning in medical image analysis. *Medical Image Analysis*. prosinac 2017.;42:60–88.
 80. Zhao R, Wang J, Yan R, Mao K. Machine health monitoring with LSTM networks. U: 2016 10th International Conference on Sensing Technology (ICST) [Internet]. Nanjing, China: IEEE; 2016 [citirano 08. ožujak 2019.]. str. 1–6. Dostupno na: <http://ieeexplore.ieee.org/document/7796266/>
 81. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. svibanj 2015.;521(7553):436–44.
 82. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. studeni 1998.;86(11):2278–324.
 83. Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, i ostali. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. U: Diamantaras K, Duch W, Iliadis LS, urednici. *Artificial Neural Networks – ICANN 2010* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010 [citirano 19. ožujak 2019.]. str. 92–101. Dostupno na: http://link.springer.com/10.1007/978-3-642-15825-4_10
 84. Suárez-Paniagua V, Segura-Bedmar I. Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction. *BMC Bioinformatics* [Internet]. lipanj 2018. [citirano 19. ožujak 2019.];19(S8). Dostupno na: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2195-1>
 85. LeCun Y, Bottou L, Orr GB, Müller K-R. Efficient BackProp. U: *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop* [Internet]. London, UK, UK: Springer-Verlag; 1998. str. 9–50. Dostupno na: <http://dl.acm.org/citation.cfm?id=645754.668382>
 86. Lisjak D. *Primjena metoda umjetne inteligencije pri izboru materijala*. [Zagreb]: Sveučilište u Zagrebu; 2004.
 87. Michelucci U. *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks* [Internet]. Berkeley, CA: Apress; 2018 [citirano 26. veljača 2019.]. Dostupno na: <http://link.springer.com/10.1007/978-1-4842-3790-8>
 88. Rasmussen CE, Williams CKI. *Gaussian processes for machine learning*. Cambridge, Mass: MIT Press; 2006. 248 str. (Adaptive computation and machine learning).
 89. Snoek J, Larochelle H, Adams RP. *Practical Bayesian Optimization of Machine Learning Algorithms*. U: *Proceedings of the 25th International Conference on Neural Information*

- Processing Systems - Volume 2 [Internet]. USA: Curran Associates Inc.; 2012. str. 2951–2959. (NIPS'12). Dostupno na: <http://dl.acm.org/citation.cfm?id=2999325.2999464>
90. Jones DR, Schonlau M. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*. 13:455–92.
 91. Wegerich SW, Wilks AD, Pipke RM. Nonparametric modeling of vibration signal features for equipment health monitoring. U: 2003 IEEE Aerospace Conference Proceedings (Cat No03TH8652) [Internet]. Big Sky, Montana, USA: IEEE; 2003 [citirano 10. prosinac 2018.]. str. 3113–21. Dostupno na: <http://ieeexplore.ieee.org/document/1234154/>
 92. Yang BS, Han T, An JL. ART–KOHONEN neural network for fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*. svibanj 2004.;18(3):645–57.
 93. Yang B-S, Han T, Hwang W-W. Fault diagnosis of rotating machinery based on multi-class support vector machines. *Journal of Mechanical Science and Technology*. ožujak 2005.;19(3):846–59.
 94. Younus AMD, Yang B-S. Intelligent fault diagnosis of rotating machinery using infrared thermal image. *Expert Systems with Applications*. veljača 2012.;39(2):2082–91.
 95. Sugumaran V, Ramachandran KI. Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mechanical Systems and Signal Processing*. srpanj 2007.;21(5):2237–47.
 96. Graney BP, Starry K. Rolling Element Bearing Analysis. *MATERIALS EVALUATION*. 2012.;8.
 97. Nyquist H. Certain Topics in Telegraph Transmission Theory. *Trans AIEE*. 1928.;47:617–44.
 98. Yang Y, Dong XJ, Peng ZK, Zhang WM, Meng G. Vibration signal analysis using parameterized time–frequency method for features extraction of varying-speed rotary machinery. *Journal of Sound and Vibration*. siječanj 2015.;335:350–66.
 99. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*. 01. listopad 1959.;148(3):574–91.
 100. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybernetics*. travanj 1980.;36(4):193–202.
 101. Bull AD. Convergence Rates of Efficient Global Optimization Algorithms. *J Mach Learn Res*. studeni 2011.;12:2879–2904.
 102. van der Maaten L, Hinton G. Visualizing non-metric similarities in multiple maps. *Machine Learning*. travanj 2012.;87(1):33–55.
 103. van der Maaten L. Barnes-Hut-SNE. arXiv:13013342 [cs, stat] [Internet]. 15. siječanj 2013. [citirano 15. srpanj 2019.]; Dostupno na: <http://arxiv.org/abs/1301.3342>

ŽIVOTOPIS

Davor Kolar rođen je 24. svibnja 1988. godine u Zagrebu. Osnovnu i srednju školu završava u Pregradi, nakon čega u akademskoj godini 2007./2008. upisuje studij strojarstva na *Fakultetu strojarstva i brodogradnje*. Diplomirao je cum laude u srpnju 2012. na smjeru *Industrijsko inženjerstvo i menadžment*.

U srpnju 2012. zapošljava se u poduzeću *Alati Stuhne* d.o.o., gdje radi na poslovima inženjera tehnologa i industrijskog inženjera. Sudjeluje u internim projektima s ciljem povećanja učinkovitosti proizvodnih procesa i kvalitete, a od 2013. vodi projekt razvoja i implementacije računalnog sustava za terminiranje proizvodnje i evidenciju rada u poduzeću. Krajem 2012. upisuje poslijediplomski studij strojarstva na *Fakultetu strojarstva i brodogradnje*, smjer *Industrijsko inženjerstvo i menadžment*. Od veljače 2014. godine zaposlen je kao asistent na *Katedri za upravljanje proizvodnjom, Zavoda za industrijsko inženjerstvo*.

Područja kojima se bavi su održavanje, razvoj i implementacija informacijskih sustava te sustavima za podršku procesima odlučivanja.

Aktivno sudjeluje u izvođenju nastave te je uključen u ostale znanstvene i stručne djelatnosti *Katedre za upravljanje proizvodnjom*. Do sada je kao autor ili koautor objavio 12 znanstvenih radova u časopisima te zbornicima radova u zemlji i inozemstvu.

Oženjen je, govori i piše engleski jezik, a služi se i njemačkim.

SHORT BIOGRAPHY

Davor Kolar was born on May 24, 1988 in Zagreb. He finished elementary and high school in Pregrada, after which he enrolled in the *Faculty of Mechanical Engineering and Naval Architecture* in the academic year 2007/2008. He graduated *cum laude* in 2012 with a master degree in *Industrial Engineering and Management*.

In July 2012, he was employed by *Alati Stuhne Ltd.*, where he worked as a technologist and industrial engineer. He participates in internal projects with the aim of increasing the efficiency of production processes and quality. Since 2013 he has been leading the project of developing and implementing a CAPP (*Computer aided production planning*) system in the company. At the end of 2012 he enrolled in a postgraduate study in mechanical engineering at the *Faculty of Mechanical Engineering and Naval Architecture*, majoring in *Industrial Engineering and Management*. Since February 2014, he has been employed as a Junior Researcher and Teaching Assistant at the Chair for Production Management at the *Department of Industrial Engineering*.

His areas of expertise are maintenance, development and implementation of information systems and decision support systems.

He is actively involved in teaching the courses and he is also involved in scientific and professional activities of the Department of Industrial Engineering. So far, as an author or co-author, he has published 12 scientific papers in journals and in proceedings in the country and abroad.

He is married, speaks and writes English and use German.