

UNIBERTSITATEKO MASTERRA:

Telekomunikazioen Teknologien Ingeniaritza

MASTER AMAIERAKO LANA

***BERPIZTEKO BIHOTZ ERRITMOAK SAILKATZEKO
IKASKETA AUTOMATIKOAN OINARRITUTAKO
ALGORITMO BATEN GARAPENA***

Ikaslea *Pitarke Ordoñez, Munia*
Zuzendaria *Irusta Zarandona, Unai*
Ikasturtea *2017-2018*

Bilbo, 2018ko Irailaren 24a

Laburpen hirueleduna

Laburpena

Dokumentu hau berpizteko bihotz erritmoak sailkatzeko teknika ez-inbaditzaileen erabileran eta garapenean oinarrituta dago; zehatzago, ikasketa automatikoan oinarritutako algoritmo baten garapena dugu helburu. Ikasketa automatikoaren erabilera, ingelesez *Machine Learning (ML)* gisa ezaguna, datu multzo batetik abiatuta ondorio esanguratsuak lortzeko aurkezten duen gaitasuna dela medio, asko nagusitu da azken urteotan. Kasu honetan, seinaleen elektrokardiogramek (EKG) aurkezten dituzten ezaugarri ezberdinez baliatuz, *Supervised learning* (Ikasketa gainbegiratua) deritzon ikasketan funtsatutako *Random Forest (RF)* izeneko sailkatzailea erabili da.

Bihotz-biriketako gaixotasunak nabarmen hazi dira mundu mailako heriotza-tasa kausarik altuenak izatera ailegatuz. Izatez, Munduko Osasun Erakundearen esanetan, herrialde garatuen heriotzen %50a eta garapen bidean dauden herrialde askoren heriotza kausa nagusietako bat gaixotasun kardiobaskularrak dira [1]. Horrenbestez, biomedikuntza esparru zientifiko honekin loturiko ikerkuntza oro sustatzeak berebiziko garrantzia hartu du. Proiektu honek, egoera hauei aurre egiteko egun existitzen diren terapiak pazienteekiko duten eragina ebaluatzeko bai eta tratamenduen kalitatea hobetzeko ezinbestekoa den EKG seinaleen analisisa eta azterketa errazteko berpizte bihotz erritmoen anotatze automatikoa ahalbidetzen du; era horretan, adituak diren medikuentzako euskarri edota laguntza izanik, diagnostiko objektibo zuzenak eta egokiak emateko ziurtasun handiagoa eskaintzen da. Ez hori bakarrik, proiektu honek etengabeko kontrolean egon behar duten pazienteen monitorizaziorako, zein larrialdi egoera baten aurrean hartu beharreko erabaki kritikoak azkarrak eta egokiak izateko ere laguntza handia ematen du.

Proiektuaren xedea erdiesteko, 857 pazientez osaturiko datu-base batekin egin da lan. Garaturiko interfaze grafikoan oinarriturik, paziente bakoitzaren seinaleen analisisa, anotazioa eta segmentazioa burutu da; eta, azkenik, 10-20 segundo bitarteko 3803 segmentuz osaturiko datu-basea lortu da. Hortik abiaturik, algoritmoaren prozesaketarako 4 eta 8 segundotako EKG seinale segmentuak erabili dira. Segmentu bakoitzeko 33 ezaugarri kalkulatu dira; eta, ondoren, ML algoritmoa elikatu da eta balioztatze gurutzatuarekin ebaluatu da. RF algoritmoak eskuragarri dituen parametro ezberdinen arabera algoritmoak sorturik, hauen errendimendua eta eraginkortasuna ebaluatzeko klase anitzeko sailkatzaileetan balio orokorra eskaintzen duten UMS (Unweighted Mean of Sensitivities) eta MulAcc (Multiway Accuracy) metrikak erabili dira. Seinale segmentu iraupen ezberdinen artean errendimendu aldeak ikusi dira, emaitza hobekak 8 segundotako seinale segmentuekin lortu direlarik. Orokorrean, sailkatzailearen zehaztasun maila %96.5koa izan da eta sentikortasun maila %88.0koa izan da, batez beste; gure emaitzek, beraz, berpizte bihotz erritmoak automatikoki identifikatzea eta sailkatzea posiblea dela baieztatu dute.

Hitz gakoak: Ikasketa automatikoa, Random Forest, balioztatze gurutzatua, berpizte bihotz erritmoen gainbegiratutako sailkapena, bihotz-biriketako gelditzea, bihotz-biriketako berpiztea.

Abstract

This document is based on the use and development of a non-invasive technique for the classification of cardiac resuscitation rhythms; specifically, its objective is the development of an algorithm based on automatic learning. In recent years, the use of *Machine Learning (ML)* has prevailed because of the ability to achieve meaningful conclusions from a set of data. In this case, a classifier based on *Supervised learning* and known as *Random Forest (RF)* has been used, taking advantage of the different characteristics offered by the electrocardiogram (ECG) signals.

Cardiovascular diseases have grown significantly becoming one of the leading causes of mortality worldwide. In fact, the World Health Organization reports that cardiovascular diseases represent half of all deaths in developed countries and that it is one of the leading causes of death in many developing countries [1]. Therefore, the promotion of research related to this biomedical scientific area is very important. This project allows automatic recording of cardiac resuscitation rhythms to facilitate the necessary study and analysis of ECG signals, with the aim of achieving improvements in the quality of treatments, as well as to pursue the evaluation of therapies on patients that are in the presence of the mentioned cardiovascular arrest situations; in this way, being a support and help for expert doctors, this project offers greater reliability on the correct and appropriate objective diagnoses that are given. Not only that, both in the continuous monitoring of patients and when making critical decisions in an emergency situation, this project offers a great help.

To achieve the project's objective, we have worked with a database of 857 patients. Based on a developed graphic interface, the analysis, annotation and segmentation of each of the signals corresponding to each patient was carried out, finally achieving a database of 3803 segments with an average duration of approximately 10-20 seconds. From there, segments of 4 and 8 seconds of ECG signal have been used for processing the algorithm. For each segment, 33 characteristics have been calculated, to then feed the ML algorithm and evaluate it with cross-validation. To evaluate the performance and efficiency of each of the algorithms created based on the different parameters offered by the RF algorithm, the UMS (Unweighted Mean of Sensitivities) and MulAcc (Multiway Accuracy) metrics have been used to offer generic values for multiple classes classifiers. In the performance obtained by the two different durations of signal segments employed, differences have been found differences, achieving more efficient results with the signal segments of 8 seconds in this case. In general, the accuracy and sensitivity average of the classifier have been 96.5% and 88.0%, respectively; our results show, therefore, that the identification and automatic classification of cardiac resuscitation rhythms is possible.

Key words: Machine learning, Random Forest, cross-validation, supervised classification of cardiac resuscitation rhythms, cardiopulmonary arrest, cardiopulmonary resuscitation.

Resumen

Este documento está basado en el uso y en el desarrollo de una técnica no invasiva para la clasificación de ritmos de resucitación cardíacos; más concretamente, tiene como objetivo el desarrollo de un algoritmo basado en el aprendizaje automático. Estos últimos años el uso del aprendizaje automático, más conocido en inglés como *Machine Learning (ML)*, ha predominado por la capacidad de conseguir conclusiones significativas a partir de un conjunto de datos. En este caso, se ha empleado el clasificador conocido como *Random Forest (RF)* basado en *Supervised Learning* (Aprendizaje supervisado) aprovechando las diferentes características que ofrecen las señales de electrocardiograma (ECG).

Las enfermedades cardiovasculares han crecido notablemente hasta el punto de ser una de las causas principales de mortalidad a nivel mundial. De hecho, la Organización Mundial de la Salud indica que las enfermedades cardiovasculares representan la mitad de todas las muertes en los países desarrollados y que es una de las principales causas de muerte en muchos países en vías de desarrollo [1]. Por consiguiente, impulsar todas las investigaciones relacionadas con esta área científica biomédica es muy importante. Este proyecto posibilita la anotación automática de ritmos de resucitación cardíacos para así facilitar el necesario estudio y análisis de las señales ECG, con el objetivo de conseguir mejoras tanto en la calidad de los tratamientos como en la evaluación de las terapias sobre pacientes ante las situaciones mencionadas de parada cardiovascular; de esa manera, siendo un apoyo y ayuda para los médicos expertos, este proyecto ofrece una mayor fiabilidad sobre los diagnósticos objetivos correctos y adecuados que se dan. No solo eso, este proyecto ofrece una gran ayuda tanto en la monitorización de pacientes con necesidad de un control continuo como a la hora de tener que tomar decisiones críticas frente a una situación de emergencia.

Para alcanzar el objetivo del proyecto se ha trabajado con una base de datos de 857 pacientes. Basándose en la interfaz gráfica desarrollada, se ha realizado el análisis, la anotación y la segmentación de cada una de las señales correspondientes a cada paciente, consiguiendo finalmente una base de datos de 3803 segmentos con una duración media de 10-20 segundos aproximadamente. Partiendo de ahí, se han empleado segmentos de señal de ECG de 4 y 8 segundos para el procesamiento del algoritmo. Por cada segmento mencionado se han calculado 33 características, para después alimentar el algoritmo de ML y evaluarlo con validación cruzada. Para evaluar el rendimiento y la eficacia de cada uno de los algoritmos creados basados en los diferentes parámetros que ofrece el algoritmo RF, se han utilizado las métricas UMS (Unweighted Mean of Sensitivities) y MulAcc (Multiway Accuracy) que ofrecen valores genéricos para clasificadores de múltiples clases. En el rendimiento obtenido por las dos diferentes duraciones de segmentos de señales empleados se han observado diferencias, consiguiendo resultados más eficientes con los segmentos de señales de 8 segundos. Por lo general, la precisión y la sensibilidad media del clasificador han sido del 96.5% y 88.0%, respectivamente; nuestros resultados demuestran, por lo tanto, que es posible la identificación y clasificación automática de los ritmos de resucitación cardíacos.

Palabras clave: Aprendizaje automático, Random Forest, validación cruzada, clasificación supervisada de ritmos de resucitación cardíacos, parada cardiopulmonar, resucitación cardiopulmonar.

Aurkibidea

Laburpen hirueleduna	2
Laburpena	2
Abstract	3
Resumen	4
Aurkibidea	5
Taulen zerrenda	7
Irudien zerrenda	9
1. Sarrera	12
2. Testuingurua	13
3. Helburuak eta irismena	15
4. Proiektuaren onurak	17
4.1 Onura teknikoak	17
4.2 Onura ekonomikoak	17
4.3 Onura sozialak	18
5. Gaiaren egoera	19
5.1 Random Forest algoritmoa	20
5.2 Aukeratutako ezaugarri bereizgarriak	22
5.3 Balioztatze gurutzatua	23
5.4 Ebaluazio irizpideak	24
6. Alternatiben analisiak	27
6.1. Programazio-lengoaia	27
6.1.1. Python	27
6.1.2. R	28
6.1.3. Matlab	29
6.1.4. Octave	30
6.1.5. C	31
6.1.6. Software plataformaren hautaketa irizpidea	31
6.1. Sailkatzaile automatikoak	33
6.1.1. K-nearest neighbors (K-NN)	33
6.1.2. Erregresio logistikoa (EL)	33

6.1.3. Support Vector Machines (SVM).....	34
6.1.4. Artificial Neuronal Networks (ANN).....	34
6.1.5. Random Forest (RF)	35
6.1.6. Sailkatzaile automatikoaren hautaketa irizpidea.....	36
7. Arriskuen analisia	37
8. Soluzioaren deskribapena	43
8.1. 1-FASEA: Bepizte bihotz seinaleen analisisian oinarritutako anotazioa eta segmentazioa.....	43
8.1.1. Kudeaketa tresnaren garapena.....	45
8.1.2. Seinaleen analisia, anotazioa eta segmentazioa.....	57
8.2. 2-FASEA: Algoritmoaren garapenean oinarritutako datu-basearen sorrera	62
8.3. 3-FASEA: Algoritmoaren garapena, ebaluaketa eta hautaketa	63
8.3.1. Seinale segmentuen prozesaketa	63
8.3.2. Ezaugarrien erauzketa	64
8.3.3. Balioztatze gurutzatua.....	65
8.3.4. Algoritmoaren garapenerako entrenamendu eta balioztatze faseak.....	67
8.3.5. Algoritmoen ebaluazioa eta hautaketa	68
8.3.6. Emaizten laburpena.....	70
9. Metodologia.....	74
9.1. Erabilitako baliabideak	74
9.2. Lan-paketeen deskribapena	74
9.2.1. Lan-taldea.....	74
9.2.2. Lan-paketeak.....	74
9.2.3. Mugarriak eta entregagaiak.....	78
9.2. Gantt diagrama	78
10. Alderdi ekonomikoak	81
10.1. Aurrekontuaren kontu-sailak	81
10.2. Aurrekontu osoaren laburpena.....	83
11. Ondorioak	84
Erreferentziak	86
I ERANSKINA: Algoritmoen emaitzak.....	92
I.I. t = 8 segundotako seinale segmentuetan oinarritutako emaitzak	92
I.II. t = 4 segundotako seinale segmentuetan oinarritutako emaitzak.....	98

Taulen zerrenda

Taula 1. Berpizte bihotz erritmoen sailkapenerako kontingentzia taula	25
Taula 2. Software plataformaren aukeraketa egiteko irizpide ezberdinen arabera ebaluaketa laburbiltzen duen taula	32
Taula 3. Sailkatzaile automatikoaren aukeraketa egiteko irizpide ezberdinen arabera ebaluaketa laburbiltzen duen taula	36
Taula 4. Probabilitatearen neurketarako irizpide taula deskribapena barne	37
Taula 5. Inpaktuaren neurketarako irizpide taula deskribapena barne.....	37
Taula 6. Pazienteen EKG seinaleen physio_ann.mat datu-base fitxategiaren erregistroen egitura	43
Taula 7. Pazienteen seinaleen inguruan informazio gehigarria ematen duen resumenDatosExt.mat fitxategiaren egitura	44
Taula 8. KDA gailutik eskuraturiko paziente bakoitzeko seinaleen informazioa gordetzen duen .mat fitxategi bakoitzaren egitura.	44
Taula 9. Paziente bakoitzeko KDA gailuak eskuraturiko datuen artean channels egiturak barnean hartzen dituen datuen balioak	45
Taula 10. Paziente bakoitzeko KDA gailuak eskuraturiko datuen artean signals egiturak barnean hartzen dituen datuen balioak	45
Taula 11. Pazienteen EKG seinaleen physio_ann.mat datu-baseko erregistroen egitura seinalearen azterketan hautaturiko seinale analisi tartearen denbora unek biltzen dituen arraya gehiturik	50
Taula 12. Pazienteen EKG seinale bakoitzeko analisi bakoitzeko informazioa biltzen duen resumenDatosExt.mat datu-baseko taularen egitura seinalearen diagnostikoaren prozesuko datu berriak atxikiturik	55
Taula 13. Sailkatzeko konplexuak diren erritmoen oinarriko arrazoiak.....	60
Taula 14. Algoritmoaren garapenerako nahitaezkoa den db_analysis.mat datu-basearen egitura	62
Taula 15. Datu-basearen berpizte bihotz erritmo klaseen proportzioa	63
Taula 16. Balioztatze gurutzatua aplikaturik lorturiko emaitzak biltzen dituen “partition.mat” fitxategiko egituraren itxura.....	66
Taula 17. “Results.mat” fitxategian biltegituraturiko emaitzen egitura	70
Taula 18. “Results.mat” fitxategian biltegituraturiko emaitzen egitura pertzentilak atziturik	70
Taula 19. Gailendutako sailkatzaileen metrikaren emaitzak	72
Taula 20. Proiektuan zehar erabilitako baliabideak.....	74
Taula 21. Proiektuan parte hartzen duen lan-taldearen deskribapena.....	74
Taula 22. Proiektuaren garapenerako lehenengo lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da	75
Taula 23. Proiektuaren garapenerako bigarren lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da	75
Taula 24. Proiektuaren garapenerako hirugarren lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da	76
Taula 25. Proiektuaren garapenerako hirugarren lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da	77
Taula 26. Lan-unitateen deskribapena.....	78

Taula 27. Proiektuaren mugarriak	78
Taula 28. Proiektuaren entregagaiak	78
Taula 29. Proiektuan parte hartzen dutenen eskulanaren prezio unitarioa.....	81
Taula 30. Aurrekontuko giza baliabideen kostuen hedapena eta jorraketa.....	81
Taula 31. Barne-orduen laburpen taula	82
Taula 32. Amortizazioen laburpen taula	82
Taula 33. Gastuen laburpen taula	82
Taula 34. Aurrekontu osoaren laburpen taula.....	83
Taula 35. 8 segundotako seinale segmentuetan oinarrituriko 175 zuhaitz eta gutxienezko behaketa kopurua 20 dituen algoritmoaren UMS eta MulAcc metriken balioen laburpen taula	98
Taula 36. 4 segundotako seinale segmentuetan oinarrituriko 75 zuhaitz eta gutxienezko behaketa kopurua 20 dituen algoritmoaren UMS eta MulAcc metriken balioen laburpen taula (t=4seg)	100

Irudien zerrenda

Irudia 1. Eraikitako kudeaketa tresna baliatuz paziente baten berpizte bihotz seinale baten erritmo kontsentsu erabakia eta segmentazioa. Kasu honetan, kontsentsuzko erritmoa asistolia, AS, izan da, eta segmentatutako seinale tarte bereizgarria ertzak horiak dituen laukizuzen gardenaren bitartez nabarmentzen da	16
Irudia 2. Random Forest algoritmoaren entrenamendu eta balioztatze faseen adibidea aurkezten da [17].	21
Irudia 3. Algoritmo baten garapenean gerta daitezkeen hiru kasu ezberdinak. Lehenengo grafikoak underfitting-a argi aurkezten du; muga lerroak ez ditu grafikoko puntu guztiak aintzakotzat hartzen. Hirugarren grafikoak, kontrara, igarritako lerroak grafikoko puntu guztiak aintzat hartzen dituela aurkezten du; hasiera batean errendimendu oneko grafikoak dela pentsa daiteke, baina ez da horrela, zarata eta ohikoak ez diren puntuak ere kontuan hartzen ditu eta. Bigarren grafikoak, igarritako lerroa nahiko ona duen grafikoak da, puntuen gehiengoa kontuan hartzen dituelarik. Algoritmoen garapenean bilatu nahi den kasua, beti, azkeneko hau da [41].....	24
Irudia 4. Datu multzoaren zatiketa Training eta Test azpi-multzoetan balioztatze gurutzatua erabilia [42].	24
Irudia 5. Proiektuan eragina duten arriskuen analisiari egokitzen zaion probabilitate-eragina matrizea.....	42
Irudia 6. Berpizte bihotz seinaleen bistaratzea, analisia eta kudeaketa egiteko gaitasuna eskaintzen duen Matlab-en funtsaturiko software plataforma tresna.....	46
Irudia 7. Datu-baseko hirugarren erregistroari dagokion pazientearen EKG eta TTI seinaleak, baita KDA gailuaren egoera mezuak eta BBB maniobraren ondoriozko konpresioak ere agertzen dira. Goiko grafikoan pazienteari dagokion EKG seinalearen lehenengo 30 segundoen bilakaera ikusten da; bertan, KDAk seinalearen azterketan zehar automatikoki aurkezten dituen mezuak ere agertzen dira, bakoitza egokitzen zaion aldiunean. Beheko grafikoak, seinale tarte horri berorri dagokion inpedantzia aurkezten du, pazienteari emandako bular-sakadekin bat.....	46
Irudia 8. Garaturiko kudeaketa tresnak gertaeratarako, analisietarako eta diagnosirako eskaintzen dituen funtzionalitate ezberdinak.....	47
Irudia 9. Gertaeretara bideraturiko lehenengo azpi-panelak eskuragarri dituen funtzionalitate ezberdinen laburpena	48
Irudia 10. Gertaeretara bideraturiko lehenengo azpi-paneleko funtzionalitateak aktibo egonik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren azterlanaren emaitza	48
Irudia 11. Analisietara bideraturiko bigarren azpi-panelak eskuragarri dituen funtzionalitate ezberdinen laburpena	50
Irudia 12. Analisietara bideraturiko bigarren azpi-paneleko funtzionalitateak aktibo egonik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren azterlanaren emaitza gehigarriak.....	51
Irudia 13. Diagnosira eta analisisira bideraturiko hirugarren azpi-panelak eskuragarri dituen funtzionalitate ezberdinen laburpena.....	52

Irudia 14. Diagnostikoetara bideraturiko hirugarren azpi-paneleko funtzionalitateak aktibo egonik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren azterlanaren emaitza gehigarriak.....	53
Irudia 15. Diagnostikoetara bideraturiko hirugarren azpi-paneleko funtzionalitateak aktibo egonik, 16.erregistroari dagokion pazientearen zortzigarren analisi tartearen azterlanaren emaitza.....	54
Irudia 16. GUIak berez eskaintzen dituen Pan, Zoom eta Data Cursor funtzionalitateen erabilera aurkezten duen 18.erregistroari dagokion pazientearen adibidea.....	56
Irudia 17. Asistolia (AS) jasaten ari den pazientearen EKG bihotz seinalea.....	58
Irudia 18. Fibrilazio bentrikular leuna (FB leuna) jasaten ari den pazientearen EKG bihotz seinalea [65]	58
Irudia 19. Erritmo antolatua (ORG) jasaten ari den pazientearen EKG bihotz seinalea [66].....	59
Irudia 20. Fibrilazio bentrikularra (FB) jasaten ari den pazientearen EKG bihotz seinalea [67]	59
Irudia 21. Anplitude handiagoa aurkezten duen fibrilazio bentrikular mardula (FB mardula) jasaten ari den pazientearen EKG bihotz seinalea. Uhinen anplitude handiagoa, tamainaz, itxuraz aldakorak, baita erritmo aldakorak ere aurkeztuz, jarduera elektriko bentrikular kaotikoa ikusten da [68]	59
Irudia 22. Takikardia bentrikularra (TB) jasaten ari den pazientearen EKG bihotz seinalea. Erritmoa erregularra, 158 taupada minutuko maiztasunarekin eta QRS konplexuekin aurkezten da [68].....	60
Irudia 23. Takikardia bentrikular geldoa (TB geldoa) jasaten ari den pazientearen EKG bihotz seinalea	60
Irudia 24. Zalantzazko erritmoa (UN) aurkezten duen 217.erregistroari dagokion pazientearen bihotz seinale tarte baten analisia. Kasu honetan, begirada bakun azkar batekin, EKG seinaleak aurkezten duen erritmoa nahasia dela esan genezake. TTI inpedantziari so eginez, ordea, interferentziak eta konpresioak aurkezten dituela ikus genezake, honek, EKG seinalean bortizki eragiten duelarik. Honekin, hortaz, seinalea garbia eta argia ez dela, eta zalantzazko erritmo gisa sailkatzea erabakitzen dela ondoriozta daiteke	60
Irudia 25. Seinaleen segmentaziora bideraturiko 265.erregistroari dagokion pazientearen seinale tarte baten analisia.....	61
Irudia 26. Corpus-aren banaketa balioztatze gurutzatuko teknika erabilita k=10 balioarekin [69]	65
Irudia 27. Soluzioaren deskribapena laburbiltzen duen proiektuaren eskema orokorra.....	73
Irudia 28. Soluzioaren deskribapena laburbiltzen duen bloke diagrama orokorra	73
Irudia 29. Proiektuaren metodologiaren Gantt diagrama	80
Irudia 30. Zuhaitz kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan (t=8seg).....	93
Irudia 31. Zuhaitz kopuruaren araberako algoritmoek aurkezten dituzten MulAcc emaitzak ehunekotan	94
Irudia 32. Behaketa kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan 125, 150, 175 eta 200 zuhaitz erabilita hurrenez hurren (t=8seg)	95
Irudia 33. Behaketa kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan 225, 250, 275 eta 300 zuhaitz erabilita hurrenez hurren (t=8seg)	96
Irudia 34. Behaketa kopuruaren araberako algoritmoek aurkezten dituzten MulAcc emaitzak ehunekotan 125, 150, 175 eta 200 zuhaitz erabilita hurrenez hurren (t=8seg)	97
Irudia 35. Zuhaitz kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan (t=4seg).....	99
Irudia 36. Behaketa kopuruaren araberako 75 eta 100 zuhaitzetan oinarrituriko algoritmoek aurkezten dituzten MulAcc emaitzak ehunekotan (t=4seg)	100

Akronimoen zerrenda

ML	Machine Learning
EKG	Elektrokardiograma
RF	Random Forest
UMS	Unweighted Mean of Sensitivities
MulAcc	Multiway Accuracy
BioRes	Bioengineering and Resuscitation
CV	Cross-validation
AS	Asistolia
ORG	Antolatutako erritmoa
FB	Fibrilazio bentrikularra
TB	Takikardia bentrikularra
AHA	American Heart Association
KDA	Kanpoko desfibrilagailu automatikoa
AI	Artificial Intelligence
OOB	Out-of-bag
CM	Confusion Matrix
Sen	Sensitivity
PPV	Positive Predictive Value
TP	True Positive
TN	True Negative
FP	False Positives
FN	False Negatives
GUI	Graphical User Interface
K-NN	K-nearest neighbors
EL	Erregresio logistikoa
SVM	Support Vector Machines
ANN	Artificial Neuronal Networks
TTI	Transthoracic Impedance
BBB	Bihotz-biriketako berpiztea

1. Sarrera

Master amaierako proiektu hau Bilboko Ingeniaritza Eskolak duen ikerkuntza sustatzailea den BioRes (Bioengineering and Resuscitation) ikerkuntza taldearen barnean garatutakoa da. BioRes taldearen ikerkuntza lana maiztasun txikiko seinaleen tratamendu digitalean oinarriturik dago, eta bereziki, seinale biomedikoen prozesamenduan. Aplikazio aldetik, taldeak garatutako metodo eta algoritmoek bihotz-biriketako geldiuenean tratamendua hobetzea eta lagungarri izatea dute helburu.

Gaur egun mundu mailan heriotza arrazoia nagusietako bat da bihotz-biriketako gelditzea. Egoera honi aurre egoteko asmoarekin, urteak dira medikuntza esparru honetan ikerketa ezberdinak sustatzen direla, betiere sistema kardiobaskularraren eta arnas sistemaren geldiuene egoeren aurrean gizakiaren bizi-aukera handitzeko helburuarekin.

Hainbat urteetan zehar, elektrokardiografiaren ikasketa bihotz seinaleen diagnostiko tresna bilakatu da. Erritmoek sufritzen dituzten aldaketa ororen analisiak terapia ezberdinen gaineko kalitateari buruzko informazio garrantzitsua eskaini dezakete, hau da, pazientearen biziraupenean eragiten duten faktore desberdinak identifikatzen laguntzen dute [2]. Hori dela eta, erritmoen sailkapen eraginkor eta egoki bat ezinbestekoa da, gaur egun adituak diren medikuek darabilten tratamendu eta teknika ezberdinen aurrean momentuko erabaki kritiko aiposenak hartzeko, baita hauen kalitatea ere, etorkizunari begira, ahalik eta gehien hobetzeko. Bihotz berpizteen erregistroen datu-baseak paziente kopuru oso handi batek osatzen dute orokorrean, ehunka edota milaka kasu eta 30-60 minutuko iraupena duten grabazioez osatuak, hauen banakako eskuzko anotazio, analisi eta interpretazioak astunak, geldoak eta ez eraginkorrak izanik; gainera, erroreak sortzeko joera handiagoa da ere.

Hori dela eta, bihotz erritmoen sistema sailkatzaile automatikoek zein erdiautomatikoek adituen datu-base masiboen eskuzko anotatze eta ikertze lana nabarmen arinduko luketela agerikoa da [3]; ez hori bakarrik, baita monitorizaturiko pazienteen kontrolean edota egoera kritikoan aurkitzen diren biktimen aurrean erabaki azkarrak eta egokiak hartzeko esanguratsuki lagunduko lukete ere. Ildo honetatik, ikasketa automatikoan oinarritutako ikerketek eta garapenek bihotz arritmiak sailkatzen dituzten eredu berriak sortzeko aukera eskaintzen dutela esan genezake; horrela, adituak diren medikuentzako euskarri izanik, diagnostiko objektibo burutsuak, egokiak eta zuzenak emateko [4].

Proiektu honek, beraz, berpizteko bihotz erritmoak sailkatzeko ikasketa automatikoan funtsatutako algoritmo baten garapena du helburu; kasu honetan, ospitaletik kanpo bihotz-biriketako geldiuena jasan duten 857 pazienteen erregistroen datu-base batean oinarrituta, gainbegiratutako ikasketan oinarritutako *Random Forest (RF)* deritzon teknika erabili da. Eredu eraginkorrago bat lortzearren, balioztatze gurutzatua (ingelesez *cross-validation, CV*) deritzon teknika eta seinale ezberdinak ezaugarritzen dituen aukeratutako 33 ezaugarri ezberdin erabili dira. Honekin guztiarekin, RF algoritmoak eskaintzen dituen parametro ezberdinen arabera sailkatzaileak sortu, aztertu, eta onena aukeratu da. Horrela, bihotz-heriotza herrialde garatuenean lehen heriotza arrazoia izanik, bihotz-biriketako geldiuenearen ondoriozko gizakiaren bizi-aukera handitzea lortu nahi da, egun erabiltzen diren tekniken, tratamenduen eta terapien eraginkortasuna handitzeko ikerketak sustatuz eta erraztuz.

2. Testuingurua

Berpizteko bihotz erritmoak 4 kategoria ezberdinetan sailkatzen dira: asistolia (AS), antolatutako erritmoa (ORG), fibrilazio bentrikularra (FB) eta takikardia bentrikularra (TB). Odolaren fluxua mantentzeko batere eraginkorrak ez diren azkeneko bi erritmoak erritmo hilgarriak dira, tratatu ezean minutu gutxitan gizakiari bat-bateko heriotza ekar diezaioketenak. Asistolia, jarduera elektrikorik eta mekanikorik aurkezten ez duen bihotz seinalea, eta antolatutako erritmoa ere, bihotz-biriketako geldiunea pairatzen ari duen pertsona baten kasuan, kontuan hartu eta tratatu beharrekoak dira. Erritmo antolatuak pultsua duten eta pultsurik gabeko erritmoetan ere azpi sailka daitezke, era horretan bost kategoria bereiziz berpizte bihotz erritmoen artean [2][5]. Pultsuaren detekzioak, ordea, EKGaz gain beste seinale batzuk behar ditu zehaztasuna hobetzeko [6]. Datu-baseak pazienteak berezkoa pultsua duen ala ez jasotzea ezinbestekoa da, eta horretarako, medikuek pazienteak tratatzen ari direnean anotazio horiek egin behar dituzte. Aurrerago azalduko denez, gure datuetan horrelako anotazioarik egitea ezinezkoa da, desfibrilagailua ez baita medikuek erabilitakoa.

Bihotz seinale batek jasaten duen erritmoaren eta bilakaeraren arabera berpizte ekintza ezberdinak aplikatu behar dira; ekintza hauen artean bihotz-biriketako berpizte (BBB) maniobrak (aireztapenak eta bular-sakadak), deskarga elektrikoak (shock) eta tratamendu farmakologikoa dira terapiatzako funtsezkoenak eta eraginkorrenak hurrenez hurren [2]. Ahalik eta etenaldi tarte laburrenako bular-sakadak eta aireztapena, bat-bateko heriotza saihesteagatik nahitaezkoak dira bihotz-biriketako edozein berpizte prozesu batean; hau da, bihotz-biriketako geldialdia pairatzen ari duen pertsona baten biziraupena BBB maniobren kalitate eta aplikazio aproposaren eta egokiaren mendekoa da proportzio handian. Horrez gain, erritmo mota guztietan aplikagarria den beste teknika ohikoetako bat desfibrilazio atalasea murrizten duen adrenalinarekin administrazioa da [7]. Biktima erritmo hilgarriaren baten aurrean egotekotan, ordea, desfibrilazioa ezinbestekoa da, erritmo hauentzako terapiarik eraginkorrena, BBB masajeaz gain, deskarga elektrikoak izaten dira eta. Pazienteak, ordea, berehalako desfibrilazioarekin erantzunik izan ezean, tratamendu farmakologikoa nahitaezkoa bilakatzen da [8]. Farmako antiarritmikoen kasuan, hiru erabilera nagusitzen dira: arritmia bentrikular kaltegarriak akabatzea, desfibrilazio elektrikoaren erraztea eta TB eta FB erritmoetara itzultzeko joera ekiditea [9]. Hauen artean, lidokaina edota amiodarona erabilienetarikoa dira; amiodaronari dagokionez, gainera, shock elektrikoaren ondoren bakarrik ez, shock-a aplikatu aurretik erabilia desfibrilazioarekiko erantzun hobea aurkezten du ere. Asistoliaren kasuan, aldiz, atropina eta epinefrina dira ohikoenak, American Heart Association (AHA) gidaren arabera hauen erabilera oso hedatua eta komenigarria ez den arren, ikerketa sakonagoak behar izanik [10].

Proiektu honetan garatutako berpizteko bihotz erritmoak sailkatzeko algoritmoaren bitartez, aipatutako terapia eta tratamendu ezberdin hauen guztien aplikazio eraginkor, fidagarri eta egokiak bilatu nahi dira, bai eta egoera kritiko baten aurrean egotekotan hauen auresate egokiak azkarragoak eta efizienteagoak izaten ere. Horretarako, 857 erregistrodun ospitaletik kanpoko berpizte hornituriko datu-base zabal batetik lortutako 3803 segmentuz osatutako datu-base batean oinarritu gara, lehenago esan bezala, ML ikasketan funtsatutako RF algoritmoa baliatuz. Algoritmo honek aurkeztu ditzakeen aplikazioen artean bi aplikazio behinen bereiziko ditugu:

- **Terapiaren doiketa erritmoa ezagututa:**

Egunerokoan aurki ditzakegun egoera ezberdinetan aplikazio honek laguntza handia ematen du. Sistema kardiobaskularraren eta arnas sistemaren geldiuonekin loturiko agertokiak ospitaleetan bertan edo, bereziki, ospitaletik kanpo gertatzen dira zoritxarrez. Larrialdi hauei aurre egiteko, dagoeneko aipaturiko teknika eta metodologia ezberdinak aplikatzea ezinbestekoa da; hots, BBB masajeak, deskarga elektrikoak eta baita drogen ezarpena ere. Momentuan aplikatu beharreko tratamendua nola eta noiz jorratu behar den zehazki ezagutu ahal izateko, biktinek momentu oro aurkezten duten erritmoa eta honen bilakaera ezagutzea ezinbestekoa suertatzen da; alegia, pazientearen egoera/erantzun harremanari buruzko zenbat eta informazio gehiago izan, aplikaturiko tratamendua orduan eta fidagarriagoa eta eraginkorragoa izateko aukera du; larrialdi egoeretan biktimaren biziraupena mantentzen lagunduko duen tratamendua gauzatzeko iraupena, aldiz, mugatua da guztiz, diagnosi oso eta egoki bat emateko denbora urria izanik. Jakina denez, bihotz-biriketako geldialdia pairatzen ari duen biktima baten lehenengo segundoak pazientearen biziraupenerako gako dira; ildo honetatik, diagnostikoan lagun dezakeen informazioa bilatzeak biktimaren biziraupenerako gako diren lehenengo segundoen galera ekar dezake, eta hortaz, honen bizi aukera jaitsi. Horrenbestez, larrialdi egoera baten aurrean berpizteko bihotz erritmoak bereizten dituen algoritmo sailkatzaile bati esker, fidagarria eta eraginkorra den diagnosi denbora-tartea murriztea posiblea da, terapia doitzuz eta berpiztearen errendimendua areagotuz.

Terapien doiketak kontrol jarraitua behar duten bihotz-biriketako gaixotasunak izateko arriskua eta joera duten monitorizatutako pazienteen kasuetan ere berebiziko garrantzia du. EKGek eskaintzen duten informazioa ez da batere argia, askotan, erritmoen artean zalantzak egonik eta, ondorioz, erritmoen aurrean akatsak sortuz; hala nola, AS eta anplitude txikiko FB erritmoak elkarren oso antzekoak dira, bien artean ezberdintzeko zailtasunak egonik. Horregatik, momentu oro kontrolean dagoen paziente baten erritmo bilakaera, adituek berehala ikus ezin ditzaketen erritmo ezaugarrietan oinarritutako algoritmo sailkatzaile batek eskaintzeak, adituen lana asko errazten dute, hauen errendimendua eta terapien kalitatea bera ere, nabarmenki hobetuz.

- **Terapiak pazienteen erritmoaren bilakaeran duen eraginaren analisia eta azterketa:**

Pazienteen berpizteen faktore zehatz eta esanguratsuak identifikatzeko bihotz-biriketako geldialdien biktimen seinaleen analisi eta azterketa sistematikoa terapia ezberdinen kalitatea ezagutzeko funtsezkoa da; hau da, identifikatutako faktoreen arabera terapia ezberdinen ikuspuntuak alderatzea posiblea egiten du, terapiak biktimaren erantzunean, hau da, bihotz seinalean duen eragina ezagutzeko gaitasuna emanaz [5].

Horrez gain, esparru honetan oinarritutako ikerketetan lan egiten duten aditu askok datu-base zabalekin lan egin behar dute, erregistroak banan-banan aztertuz eta anotatuz; horrek, nahitaezkoa den prozesu hau neketsua, geldoa eta eraginkorra ez izatea eragiten du. Izan ere, iraupen luzeko pazienteen berpizte seinaleen anotazioek atzera begirako berpizte datuen azterketa eta analisi aberatsago bat izatera laguntzen dute, baita pazienteari aplikaturiko terapiaren eta erantzunaren arteko elkarrekintza hobeto ezagutzen ere [3].

Bihotz erritmoak sailkatzeko algoritmo automatikoek, beraz, ikerketa hauentzako ezinbestekoa den prozesu hau asko errazten dute, eramangarriagoa bilakaturik, eta ondorioz, terapien, tekniken eta metodoen eraginkortasuna eta kalitatea areagotzeko ikerketak sustatuz.

3. Helburuak eta irismena

Proiektu honek sistema kardiobaskularrarekin eta arnas sistemarekin loturiko terapiak tratatzeko berpizte gidek gomendatzen dituzten teknika eta metodo ezberdinen eraginkortasuna eta fidagarritasuna areagotzeko, baita hauek hobera egitea bilatzen eta ahalbidetzen duten ikerketak errazteko eta sustatzeko ere, berpizteko bihotz erritmoak identifikatzen eta sailkatzen dituen algoritmo automatiko baten garapena du xede.

Aipatutako helburu nagusi hau lortu ahal izateko jarraian azalduko diren hiru azpi-helburuak betetzea ezinbestekoak dira:

1. Berpizte bihotz seinaleen analisisian oinarritutako anotazioa eta segmentazioa

Proiektuak duen helburu nagusia erdiesteko, berpizte prozesuan zeharreko bihotz seinaleen anotazioak berebiziko garrantzia dute. Horrenbestez, jorratu beharreko lehendabiziko pausua ospitaletik kanpoko 857 paziente ezberdinen bihotz erregistroekin osatutako datu-base baten analisi sakona, erritmoen anotazioa eta segmentazioa da.

Honakoa ahalbidetzeko, erregistro bakoitzaren azterketa eta anotazioa egiteko aukera ematen duen tresna baten garapena beharrezkoa da. Kasu honetan, garatutako software plataformak, pazienteen EKG seinaleez gain, seinale bakoitzaren inpedantzia ere erakusten du, erritmoen anotazioetarako honek eskaintzen duen informazioa, pazienteen momentu bakoitzean bular-sakadak jasotzen dituen ala ez adibidez, baliagarria eta lagungarria baita. Horrez gain, benetako erritmoaren anotazioa kontsentsu bidez erabakitzea ahalbidetuko duen sei aditu ezberdinen erritmoen anotazioak ere eskuragarri ditu.

Ospitaletik kanpoko berpizte bihotz seinaleez osatutako datu-basean oinarritutako kudeaketa tresna eskuragarri izanik, pazienteen banakako seinaleen anotazioa eta segmentazioa izan da hurrengo pausua. Hau da, seinale bakoitzak dituen medikuen anotazio une bakoitzeko, kontsentsuzko erritmo batera ailegatu eta erritmo hori den ezaugarritzen duen seinale tarte esanguratsuenak aukeratu (segmentazioa). Segmentazio eta anotazio prozesua Irudia 1ean erakusten da.

2. Datu-base baten sorrera

Garatutako algoritmoaren eraginkortasuna eta efizientzia baieztatzeko, ahalik eta paziente gehienaren seinale segmentuak biltzen dituen datu-base baten sorrera nahitaezkoa da. Datu-baseak ikasketa automatikoan oinarritutako algoritmo baten garapenerako beharrezkoa den informazioa eta datuak biltzen ditu. Kasu honetan, aurreko pausuan lortutako paziente guztien anotatutako seinale segmentuak biltzen dituen datu-basea lortzen da, guztira 3803 analisi esanguratsuekin osaturikoa, lagungarria den informazio eta datu gehiagorekin batera.

3. Algoritmoaren garapena, ebaluaketa eta hautaketa

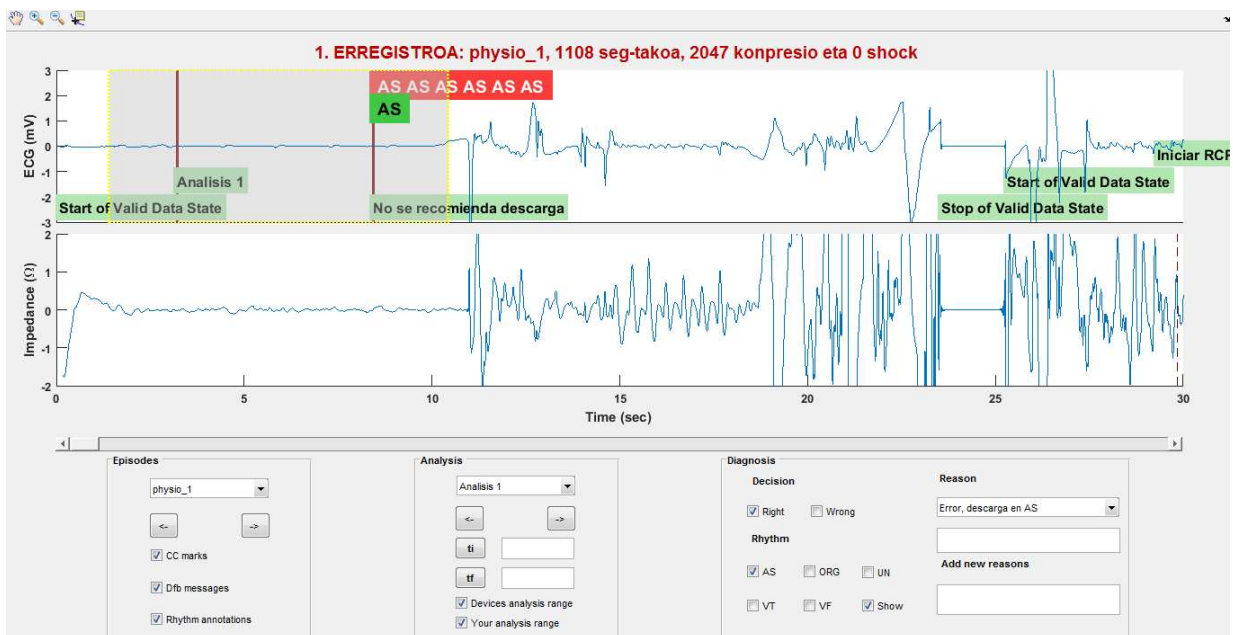
Aurretik azaldu denez, proiektuaren helburu nagusia bihotz-biriketako berpizte egoeren aurrean aplikatu beharreko metodo eta teknika ezberdinen eraginkortasuna, fidagarritasuna eta kalitatea

hobetzen duten berpizte erritmoen sailkatzaile automatiko bat sortzea da, horretarako ikasketa automatikoan oinarritutako algoritmo baten garapena ezinbestekoa izanik.

Sortu nahi den algoritmo batean entrenamendurako dauden datuen kopurua urria izatekotan, algoritmoak ezagutza orokortzeko zailtasunak izango ditu underfitting-a eraginez. Askotan, kontrara, ereduak gertaera partikular batzuk bakarrik ikastera doitzen dira, horrela, overfitting-a sortuz, eta datu berriak ezagutzeko zailtasunak edo ezintasuna izanik. Horregatik, arazo hauei aurre egiteko asmoarekin, algoritmoaren garapenerako balioztatze gurutzatua erabiliko da; hau da, algoritmoa garatzeko beharrezkoa den datu multzoa bi azpi-multzotan banatuko da: entrenamendurako eta ereduak ezagutuko ez dituen balioztatze prozesurako, hots, testeatzeko. Horrela, sorturiko eredu ahalik eta eraginkorra izatea lortu nahi da.

Horrez gain, algoritmo sailkatzaile fidagarriak lortzeko, aztertzen diren elementuen inguruko informazioa eskaintzen duten, diskriminatzaileak diren eta independenteak diren ezaugarrien hautaketa beharrezkoa da ere, kasu honetan, seinaleak ezaugarritzen dituzten 33 ezaugarri ezberdin erabiliz.

Seinaleen ezaugarri bereizgarriak lorturik, baita balioztatze gurutzatua aplikaturik ere, ikasketa automatikoan oinarritutako gainbegiratze ikasketan funtsatutako RF teknika erabiliko da, parametro ezberdinen araberako ereduak lortzea ahalbidetzen duena. Azkenik, lortutako sailkatzaileen errendimendua ebaluatzeko, *Multiway Accuracy (MulAcc)* eta *Unweighted Mean of Sensitivities (UMS)* balioak dira kontuan hartuko direnak. MulAcc parametroak sailkatzaileak bere osotasunean aurkezten duen doiketa neurtzen du; hau, ordea, klaseen desorekari dagokionez, sentikorra da oso. UMSak, kontrara, sailkatu nahi diren erritmo klase guztiak pisu berdinarekin ebaluatzen ditu, klaseen arteko desorekak kalterik sor ez dezan [2]. Horrela, lortutako erritmoen sailkatzaile eredu guztien arteko emaitzen ebaluaketa sakona eginik, eraginkortasun eta fidagarritasun maila handiena eta onena aurkezten duen sailkatzailea hautatzea posiblea da, xede nagusia arrakastatsu erdietsiz.



Irudia 1. Eraikitako kudeaketa tresna baliatuz paziente baten berpizte bihotz seinale baten erritmo kontsentsu erabakia eta segmentazioa. Kasu honetan, kontsentsuzko erritmoa asistolia, AS, izan da, eta segmentatutako seinale tarte bereizgarria ertzak horiak dituen laukizuzen gardenaren bitartez nabarmentzen da

4. Proiektuaren onurak

Proiektu honen bitartez etorkizuneko gizartean onura teknikoak, ekonomikoak eta sozialak lortzea espero da. Jarraian, esparru bakoitzean eskaini ditzakeen ekarpenak sakonago azalduko dira:

4.1 Onura teknikoak

Onura teknikoak dagokienez, dagoeneko aipatutakoa azpimarratuko da nagusiki:

Alde batetik, proiektu honek aurkezten duen ekarpen teknikorik esanguratsuena, agertoki ezberdinetako sistema kardiobaskularrarekin eta arnas sistemarekin loturiko terapiak tratatzeko egun erabiltzen diren teknika eta metodoen eraginkortasuna eta errendimendua handitzeko eta hobetzeko, baita ekarpen berriak egiteko ere, xedea duten ikerketak sustatzeko berpizte erritmoen sailkatzaile automatiko baten garapena da.

Aipatuenez, biomedikuntza esparruan oinarrituriko ikerketa askotan bihotz berpizte seinaleen analisi, anotazio eta identifikazio doituak eta zehatzak egiteak berebiziko garrantzia du; izan ere, atzera begirako berpizte datuen azterketa aberatsago bat eta pazienteari aplikaturiko terapiaren eta erantzunaren arteko elkarrekintza hobe bat eskaintzen dute. Arritmien identifikazioa eta sailkapena normalean, ordea, prozedura konplexua da, zehaztasun maila handia lortzeko zailtasunak egonik eta askotan erritmoen anotazioetan akatsak egiteko aukera handiagoa izanik ere. Orokorrean, gainera, Holter monitore zein KDA baten bitartez bildutako ehunka EKG erregistrok osatzen dute datu-basea, adituentzako bihotz seinaleen azterketa oso astuna den prozesua bilakatuz. Horregatik, ikerketa hauek guztiak sustatzeko modu eraginkor bat bihotz berpizte sailkatzaile automatiko baten erabilera da, ikertzaileen lana eramangarriagoa eta errazagoa izateko.

Ikerketen sustapenari laguntza handia eskaintzeaz gain, etengabeko kontrolean dauden pazienteen monitorizazioetan, baita egoera kritikoan aurkitzen diren biktimen egoeretan ere, berpizte bihotz erritmoen analisia garrantzi handikoa da, tratamendua behar duten erritmoen bilakaera aurreratzeko oso lagungarriak baitira. Larrialdi agertokietan, gainera, pazienteak tratatzeko denbora oso urria da, hauei laguntzeko beharrezkoak diren teknika, metodo eta terapia egokien hautapena eta erabilpena berehalakoak izan behar delarik. Hori dela medio, erritmoen sailkapena automatikoa izateak prozesu hau asko azkartzen du, terapiak eraginkorragoak eta errendimendu handiagokoak izanik.

4.2 Onura ekonomikoak

Proiektu honek xede duen bihotz berpizte seinaleen sailkatzaile automatikoaren garapenak, alde batetik, momentu oro kontrolean dauden pazienteen monitorizaziora eta larrialdi egoetara bideraturik dauden ikerketak sustatzen laguntzen du. Bestetik, aipatutako egoeretan maneiatzen eta aplikatzen diren terapien, metodoen eta tekniken aurreraketa eta aplikazioa ere ahalik eta eraginkorrenak eta fidagarrienak izateko baliagarria da. Proiektuan garatutako bihotz erritmoen sailkatzaile automatikoa, beraz, egun erabiltzen diren tresna ezberdinetan aplikagarria da, baita

etorkizun hurbilean sor daitezkeen tresna berrietan ere, hauen sendotasuna eta errendimendua areagotuz.

Horrez gain, egun larrialdi egoeretan oso erabilia den ekipoetako bat KDA da, erritmoak aurreratzeko eta deskarga emateko gomendio automatikoa ematen duen algoritmoa barne duena. Pazientearen erritmoaren egoera/bilakaera eta terapien aurrean aurkezten duten emaitzen arteko elkarrekintzaren deskribapen xehe bat lortzeko, ordea, gailu hauek darabilten algoritmoa ez da nahikoa, eta hortaz, erritmoen sailkatzaile zehatzago baten beharra sortzen da. Proiektu honek garatzeko helburu duen berpizte bihotz erritmoen sailkatzaile automatikoa, aurreratzeko ahalmen fidagarriago bat lortzearren etorkizun hurbileko KDA ekipoetan ezartzea ekar dezake, KDA gailuen salmenta haziz eta indartuz.

Honekin guztiarekin, lan honen ekarpen nagusia bihotz-biriketako berpiztearen alorrera bideraturik egoteagatik, medikuntza esparruarekin loturiko merkatua hazi, zabaltu eta indartu daitekeela esan nahi da. Osasun publikoari dagokionez, garatzen den erritmoen sailkatzaile automatikoaren erabilpena gizakiaren bizi-aukera eta bizi-kalitatea handitzeko eta hobetzeko asmoz egiten diren ikerketetara eta egunerokoan eman daitezkeen larrialdi edota monitorizazio egoeretara bideraturik dago. Gainera, hil zorian dagoen gaixoa ahalik eta azkarren berpizteak, biktimak izan ditzakeen ondorio neurologikoak ekidin ditzake, horrela, osasun publikoak pertsona hauek sendatzeko inbertitu beharreko diru kopurua txikiagoa izanik.

4.3 Onura sozialak

Aurretik aipatu denez, bihotz-biriketako medikuntzaren alorrean berpizte bihotz erritmoen identifikazio eta sailkapen zehatz eta doi baten beharra dago. Bihotz-biriketako geldialdia pairatzen duen paziente batean momentuko erritmoen bilakaera aurrean edota ezaguturik, erritmo hauek biktimarengan sor ditzaketen ondorioak aurretik ezagutu daitezke, nahitaezkoak diren ekintzak eta terapiak aplikatuz. Horrez gain, algoritmo automatiko honen garapenak sustatzen dituen ikerketen ekarpenen artean, sistema kardiobaskularraren esparruarekin loturiko edozein hobekuntza eta berrikuntza dago.

Onura sozialei dagokienez, beraz, sistema kardiobaskularraren egoeren aurreko biziraupen-tasaren hazkuntza dakarrela agerikoa da; monitorizazio zein larrialdi egoeren aurrean, edota ikerketetan arrakasta lortuz, pazienteen biziraupen-tasa asko haziko da, alor honen inguruko munduan aurrerakuntza handiak erdietsiz.

5. Gaiaren egoera

Ikasketa automatikoa ereduaren ezagutzaren, probabilitate teoriaren, optimizazioaren eta estatistiken bilakaerak sortutako *Artificial Intelligence (AI)* esparruaren barnean gailendutako adarretako bat da. Helburu nagusia konputagailu programek datuetan eta esperientzian oinarriturik ikastea, erabilera ereduak ezagutzeko algoritmoak eraikitzea, eta edozein arazo baten aurrean erabaki adimentsuak azkar hartzea da.

Azken urteetan Machine Learning (ML) ikasketa automatikoa ikerketa esparruan aktibo egon da; honen xede nagusia aplikazio ezberdinetan arrakasta lortzea da, medikuntza alorrean agertzen diren arazoei aurre egitea aplikazioetako bat izanik. Ikasketa automatikoak, normalean geldoak eta subjektiboak izaten diren adituek eskuz bideraturiko prozesuak automatizatzeko gaitasuna eskaintzen du. Hori dela eta, eta aurretik aipatu izanenez, MLaren erabilerak adituen ikerketa denbora aurreztu dezakete, emaitza inpartzialak eta objektiboagoak ere aurkeztuz [11]. Gainera, medikuntzan tamaina handiko datu-baseekin lan egiten da, datuen etengabeko aldaketak eta gizakiek eskuz analisiak egitearen ondorioz sortzen diren erroreak ohikoak bilakatzen direlarik [12]. Horregatik, interpretazio eta analisi klinikoetarako ikasketa automatikoan oinarritutako tekniken erabilera ezinbestekoa da.

Baldintza patologikoen identifikazio eta interpretazioak, aztertzen ari den alorrean esperientzia handia duten aditu kopuru handi baten beharra du. Zenbaitetan, ordea, adituen kopurua ez da nahikoa, eta adituen arteko desadostasunak agertzea ere gerta daiteke. Hori dela medio, esperientzia eta kontzentrazio maila handia behar duten, baina erroreak gauzatzeko joera duten adituen beharra ezinbestekoa den medikuntza eremuetan, ikasketa automatikoan funtsaturiko teknika eta metodoek diagnostikoen zehaztasuna eta medikuen arteko estandarizazioa hobetu dezakete; ez hori bakarrik, konputagailuetan oinarritutako erabakiak hartzeko tresnen garapenean ere hobekuntzak eskain ditzakete [11].

Ikasketa automatikoa medikuntza esparruan aplikagarriak diren teknika eta metodo multzo handi batek osatzen dute. Sailkatzaile automatikoak ezagunenetarikoak dira, egunerokoan aurki daitezkeen agertoki askorekin bat etortzen baitira. Gaur egun, herrialde garatueneren heriotza kausa nagusia gaixotasun kardiobaskularrak diren mundu honetan, EKGak funtsezko garrantzia du; izan ere, medikuek seinale hauetan irregulartasunak identifikatuz, bihotzak pairatu ditzakeen anomalia desberdinetaz jabetu daitezke. Hori dela eta, elektrokardiogramak eskaintzen duen erabilera kliniko zabalak eguneroko monitorizazio zein analisisetan, baita patologia anizkoitzen diagnosian ere, erabilgarria eta baliagarria izatea posiblea egiten du [13]. Ondorioz, prozesu hauen guztien zehaztasuna, abiadura, fidagarritasuna eta errendimendua areagotzeko eta errazteko, EKG seinaleetan oinarritutako erritmoen sailkatzaile automatikoak beharrezkoak dira.

Azkeneko hamarkadetan, EKG erregistroen analisi automatiko baterako metodoen garapenerako dedikatutako ahalegina handia izan da, bihotz taupada mota ezberdinen artean ezberdintzeko, baita arritmiak identifikatzeko eta sailkatzeko ere, ML ikasketan oinarritutako algoritmo asko burutu direlarik. Algoritmo hauek erabilitako sailkapen eskemaren eta EKG seinaleetatik ateratako ezaugarrien arabera asko aldatzen dira. Kardiologo batek, diagnostiko batera ailegatzeko beharrezkoa

den elektrokardiograma baten azterketa egitean, morfologia eta denbora domeinuko ezaugarriak oinarri hartzen ditu; hala nola, uhinen eta tarteen iraupena, erritmoaren erregulartasun/irregulartasuna, P uhinaren eta QRS konplexuaren arteko elkarrekotasuna, anplitudea, polaritatea eta abar. Aipaturiko ezaugarri hauek guztiak bihotz seinaleen elektrokardiogramen analisi automatikoan naturalki miatu egiten dira; baliabide konputazionalak, ordea, informazio gehiago eta begi-bistaz lortzeko zailagoak edota ezinezkoak diren ezaugarriak eskuratzeko aukera ematen dute. Horregatik, urteetan oinarri hartu den ML ikasketa automatikoari esker, EKGak eskain ditzakeen parametro estatistiko, frekuentzia domeinuko ezaugarri edota konplexuagoa den informazioan funtsaturik sailkatzaile automatiko eraginkorrak lortu izan dira. Sailkapen eskemei dagokienez, ikertzaile batzuek metodo sinpleetan oinarritutako algoritmoak aukeran nahiago dute, beste batzuek konplexuagoak diren algoritmoetan oinarritutako sailkatzaileekin lan egitea nahiago dutelarik [14]. Medikuntza esparruko arazoetara aplikagarriak diren teknika sailkatzaileen artean Neural Networks, Support Vector Machines, Decision Trees, Logistic Regression, Bayesian Learners eta Random Forest algoritmoak daude.

Proiektu honetan bihotz seinaleen EKGetatik atzituriko lau motatako ezaugarri ezberdinetan oinarrituriko ezaugarriak erabili dira: denbora domeinuan oinarriturikoak, ezaugarri espektralak, denbora-frekuentzia ezaugarriak eta konplexutasun ezaugarriak. Horrez gain, algoritmoak garatzerakoan *overfitting* eta *underfitting* ondorioei aurre egiteko balioztatze gurutzatua aplikatu da. Berpizte bihotz erritmoen sailkatzaile automatikoari dagokionez, Random Forest algoritmoa hartu da oinarri, honek aurkezten dituen parametro ezberdinen arabera sailkatzaile ezberdinen errendimendua konparatzeko aukera eskaintzen duelarik. Azkenik, lortutako emaitzen ebaluazioa egiteko MulAcc eta UMS balioak dira irizpide gisa hartu direnak. Algoritmoaren garapenerako eta emaitzen ebaluaziorako jarraituriko metodologia azaldu aurretik, hauei loturiko oinarritzko eta beharrezko kontzeptu teorikoak errepatatuko dira jarraian.

5.1 Random Forest algoritmoa

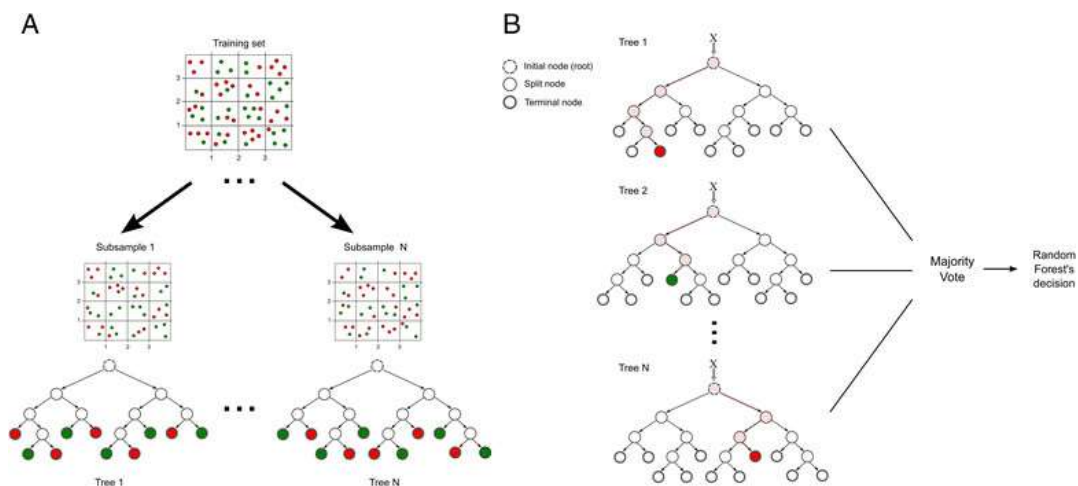
Erabaki zuhaitzak (ingelesez *Decision Tree, DT*) [15], gainbegiratutako sailkatzaile metodoen artean erabilienetarikoa da, sendoak, interpretatzen errazak eta parametrikoak ez diren metodoak baitira. Aldagaien espazioan ondoz ondoko partizioak eginez, partizio ordezkarien homogeneotasuna maximizatzen duten atalase balioa eta aldagaia bilatzen funtzionatzen dute [16]. Hau da, DT determinista eta sinplea den datuen egitura bat besterik ez da, sailkapen problema zehatz bati aurre egiteko erabaki arauen eredua burutu eta jarraitu behar duena. Nodo bakoitzean, banaketa erabaki batera heltzeko zuhaitzak, hau eraikitzen den eraren arabera, ezaugarri bat edo batzuk aukeratzen ditu, eta hosto nodoak datu optimoak dituen zatitzeari uzten dio; hots, partizioak zatiketa guztiak guztiz homogeneoak diren arte jarraitzen du. Zuhaitzaren hosto nodo horrek azkeneko emaitza bat eskaintzen du, sailkapenen kasuan klase bakoitzaren probabilitatea erakutsiz. DT baten eraginkortasunerako faktorerik erabakigarriena nodoen zatiketa prozesuaren efizientzian oinarri hartzen du, zatiketa egiteko metodorik ezagunenak Gini Ezpurutasuna (ingelesez *Gini Impurity, GI*) eta Informazio Irabazia (ingelesez *Information Gain, IG*) direlarik hurrenez hurren [17].

DT zuhaitzen arazo nagusietako bat, ordea, entrenamendu datuekiko *overfitting*-a da. Mugak ezarri ezean, zuhaitzaren entrenamendu faseak entrenamendu datuetara moldatzeko joera du, entrenamendu laginekiko ezberdinak diren datu ezberdinetarako iragarpena sentikorra izatea eragiten

duelarik. Horren ondorioz, errore mota hauek zuzendu eta zuhaitzen doitasuna nabarmenki hobetzen duten RF basoak sortu ziren [18].

RF sailkatzaile automatikoa, ikasketa automatikoan oinarritutako gainbegiratze ikasketan funtsatutako Leo Breimanek garatutako algoritmoa da [19]. Teknika hau, DT zuhaitzen, hau da, sailkatzaile ahulen multzo batetik abiatu datu kopuru handi batekin eta aldagai anitzekin lan egiteko guztiz aproposa da, zuhaitz bakoitza lantzeko orduan lagin ezberdinak erabiltzen baitira; horrela, eredu eraikitzeke aldagai eta datu guztien erabilera bermatzen da [20].

Aipatuenez, RF ikasketa automatikoa DT zuhaitzez egituraturiko sailkatzaile multzo batean oinarritzen da. Sailkatzaile honek oso indartsuak, ahaltuak eta eraginkorrak diren bi ML teknikez baliatu egiten da: *Bagging* prozesuaz [21] eta ausazko ezaugarrien hautaketaz [22]. RF algoritmoaren ideia nagusia, *Bagging* estrategiaren bitartez ausaz aukeratutako laginetatik ausaz aukeratutako ezaugarri batzuetan oinarritzen den sailkatzaileak diren zuhaitz ezberdinen multzo bat eraikitzea da, ondoren, klase etiketa bat lortzeko, aipaturiko zuhaitzek emandako sarrera bektore batengatik botoa emanaz [23]. *Bagging* prozesuan, beraz, zuhaitzetako bakoitza entrenamendu datuetako *bootstrap* laginekin [24] entrenatzen da eta iragarpenak zuhaitzen botoen gehiengoak erabakitzen du [22]. RF basoa banaketa berdina bermatzen duten, independenteki lagindutako ausazko bektore baten balioen araberakoak diren eta partizio errekursiboa jarraitzen duten DT zuhaitz bitar multzo handi batek osatzen du. Zuhaitz bitar hauetako bakoitza eraikitzeke, entrenamendu datuetatik N objektu dituen *bootstrap* lagin bat erauzi egiten da, ondoren, errekursiboki ezaugarrien azpimultzo bat aukeratuz eta hautatutako ezaugarriarik egokienarekin nodo borte bakoitza bigarren mailako bi nodoetan banatuz, zuhaitza *bootstrap* laginetara doitzen delarik [23]. RF algoritmoaren iragarpen errendimendua balioztatzeke, *out-of-bag (OOB)* laginak erabiliz entrenamendu fasearekiko paraleloan balioztatze-gurutzatua deritzon ebaluazio metodoa gauzatzen da. Esanenez, entrenamendu fasean, zuhaitz bakoitza *bootstrap* lagin partikular batekin hazi egiten da. *Bootstrapping*-a entrenamendu datuen ordezkapen laginketaenez, sekuentziatako batzuk laginetik kentzen dira, beste batzuk laginean errepikatzen direlarik. Baztertutako sekuentzia multzo hauek OOB lagina osatzen dute hain zuzen ere [22]. Hurrengo irudiak, aipatutako RF algoritmo baten entrenamendu eta balioztatze faseak aurkezten ditu:



Irudia 2. Random Forest algoritmoaren entrenamendu eta balioztatze faseen adibidea aurkezten da [17].

RF algoritmoa parametro ezberdinen arabera sailkatzaile eredu ezberdinak garatzea ahalbidetzen du. Kasu honetan, zuhaitz kopuruaren eta zuhaitz bakoitzak duen nodo kopuruaren araberrako sailkatzaile eredu desberdinak aztertu dira.

Zuhaitz kopurua:

Basoa osatzen duten zuhaitz kopuruak sailkatzailearen iragarpenaren zehaztasunean eragina du. Teorikoki, algoritmo honek erabiltzen duen *Bagging* teknika, soluziorik onena aurkitzen oso eraginkorra da; hortaz, balio zehatz batetik aurrera zuhaitzen kopurua handitzeak, pixkanakako doitasunean igoera izatea eta koste konputazional handiagoa izatea eragin beharko luke [25].

Zuhaitz bakoitzaren nodo kopurua:

Zuhaitzen sakontasuna mugatzen duen parametroaren balioa zehaztea ere garrantzitsua da RF algoritmoan; hau da, erabakiak egiteko orduan noiz gelditu behar den jakiteko, adar bakoitza sortzeko beharrezko kasuak zehaztea. Horrela, parametro honi zehazturiko balioa bost izanik, zuhaitzaren adarretako batek erabaki batera ailegatuta lau kasu besterik erabili behar izan ez dituela ikusten badu, ez du beste erabaki adar bat sortuko.

Algoritmo honek MLen beste algoritmo batzuekin alderaturik aurkezten dituen abantailen artean hurrengoak daude: alde batetik, beste metodo batzuekin konparaturik lortzen den zehaztasuna handiagoa da. Gainera, datu kopuru handiak biltzen dituzten datu-base zabalekin efizientea da oso, eta sartzen duen ausazkotasuna medio, *overfitting*-a ekiditen du. Gainera, RF algoritmoak aurkezten duen abantaila berezko bat, analizatzen diren elementuak deskribatzen dituzten ezaugarri bereizgarri garrantzitsuenak aukeratzeko gaitasuna da. Horrez gain, gure kasua den klase anitzeko problemetan oso erraz aplikagarria dela ere aipatzekoa da.

5.2 Aukeratutako ezaugarri bereizgarriak

Sailkatzaile automatikoa garatzeko prozesuan beharrezkoak diren datu-baseko datuak deskribatzen dituzten elementuak dira, sortzen diren algoritmo ereduaren eraginkortasun eta errendimendurako ezaugarri hauen hautaketa pausu erabakigarria izanik.

Proiektu honetan lortutako 3803 EKG seinale segmentu bakoitzeko 33 ezaugarri bereizgarri kalkulatu dira. Bihotz seinale segmentu bakoitzetik erauzitako ezaugarri hauetako bakoitzak berpizte bihotz erritmo bakoitzak bere aldetik aurkezten dituen bereizitasunak nagusitzen dituzte, hauen deskribapena ahalik eta zehatzena izan daiten. Aukeratutako eta aipatutako ezaugarri multzoa 4 talde ezberdinetan banatu daitezke [26]:

- **Denbora domeinuko ezaugarriak:** anplitudea, malda, laginen banaketa edo bihotz-erritmoa deskribatzen dute. Definitutako ezaugarri hauen artean hurrengoak erabili dira: atalasea zeharkatzeko tartea (*Threshold Crossing Interval, TCI*) [27], atalasea zeharkatzeko laginen kopurua (*Threshold Crossing Sample Count, TCSC*) [28], esponentzial estandarra (*Standard Exponential, Exp*) [28], eraldatutako esponentziala (*Modified Exponential, Expmod*) [29], balio absolutua (*Mean Absolute Value, MAV*) [30], *count1*, *count2* and *count3* [31], *x1* eta *x2* [32], *bCP* [33], EKG seinalean dauden QRS konplexu kopurua (*Number of Peaks, Npeak*), QRS seinaleen anplitudeen arteko aldean desbideratze estandarra (*Standard Deviation of Amplitude Peak, stdAmpPeak*) eta bihotz seinaleen ondoz ondoko R uhinen arteko distantzien desbideratze estandarra (*Standard Deviation of RR, stdRR*).

- **Ezaugarri espektralak:** espekro-kontzentrazioa, momentu espektral normalizatuak edo maiztasun-banda ezberdinetan potentzia edukiera erlatiboak kuantifikatzeko erabiliak bereziki. Hauen artean hurrengoez baliatu gara proiektuan: FB iragazkia (*vFleak*) [34], *M*, *A1*, *A2* eta *A3* [34], *x3*, *x4* eta *x5* [35], *bWT* [33] eta *bW* [33].
- **Denbora-maiztasun ezaugarriak:** hemen erabilitako ezaugarri bakarra, EKG seinalearen *wavelet* analisisan oinarritutako *Li* [36] ezaugarri bereizgarria izan da.
- **Konplexutasun ezaugarriak:** EKG bihotz seinalearen konplexutasunaren neurri adierazgarrienak eta esanguratsuenak: konplexutasun neurria (*Complexity Measure, CM*) [37], kobariantza (*Covariance, CVbin*), EKG seinaletik erauzitako seinale bitarraren azalera (*abin*), maiztasuna (*Fraqbin*), eta *Kurtosis (Kurt)* [38], fase espazioaren berreraikitzea (*Phase Space Reconstruction, PSR*) [38], Hilbert transformatua (*Hilbert transform, HILB*) [39] eta entropia lagina (*Sample Entropy, SamEn*) [40].

5.3 Balioztatze gurutzatua

Datu klase batzuk iragarri eta sailkatzeko datu multzo batekin lan egiten denean, entrenamendu eta balioztatze fase baten bitartez ahalik eta doitasunik zehatzena eskaintzen duen sailkatzaile eredua lortzea xede nagusia da. Ereduaren lehenengo bertsioan lortutako zehaztasuna ona izatekotan, datuen ezaugarri bereizgarrien kopurua handitzen edo txikitzen, hauen aukeraketa aldatzen edota ikasketa automatikoaren ML eredian ezaugarrien ingeniariak aplikatuz, aurretiaz lortutako doitasun maila handitzen saiatzen da. Hala ere, ereduak beti emaitza onak ez ematea posiblea da.

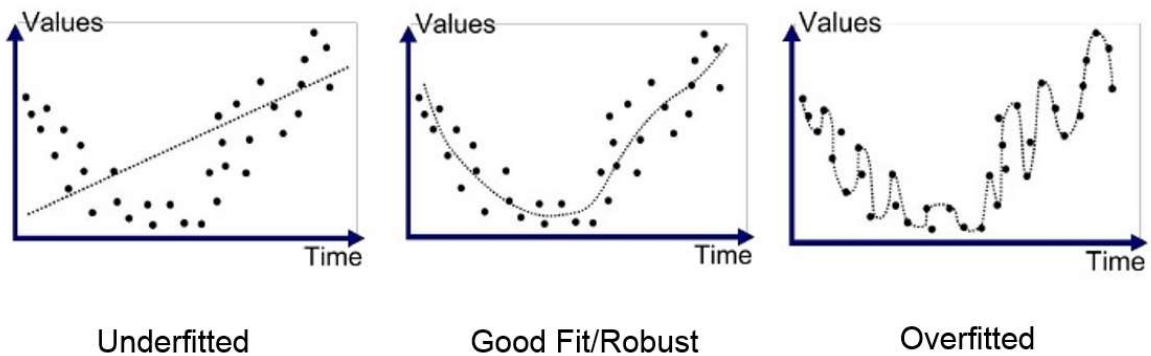
Jakina denez, sailkatzaile automatiko baten garapenerako bereizi eta sailkatu nahi diren laginen datu-base bat ezinbestekoa da. Aipatutako datu-basean biltegitutako lagin multzoa sailkatzailea entrenatzeko, baita aurrerago sailkatzailea testeatzeko ere, erabiliz gero, sortutako algoritmo ereduaren igartze-tasa oso ona izango da, doitasun maila oso altua lortuz. Ildo honetatik, agertuko litzatekeen arazorik nagusia, algoritmoak datu berriak orokortzeko gaitasunik edukiko ez lukeela litzateke; hots, sailkatzailearen entrenamendurako erabilitako datuak ez diren lagin berriak erabilia, algoritmoa datu berri horiek sailkatzeko ezintasunean edota zailtasunekin aurkituko litzatekeela, dagoeneko ezagutzen dituen datuak soilik ondo sailkatuz. Azalduko arazo honi overfitting-a deritza. Horrez gain, algoritmoa garatzeko eskuragarri dagoen datu-basea txikia izatekotan, hau da, algoritmoa sortzeko eskuragarri dauden datu kopurua urria izatekotan, algoritmoari ezagutza orokortzeko zailtasunak izatea gerta dakioke, underfitting-a sortuz. Irudia 3-ak aipatutako bi kasu hauek kasu idealarekin bat alderatzen ditu.

Aipatutako overfitting arazoa ekiditeko ezagutzen diren estrategia ezberdinen artean, *k* partiziodun balioztatze gurutzatua (*k-fold cross validation, k-fold CV*) deritzon teknika da aplikatu den estrategia. *K* partiziodun balioztatze gurutzatuan datu-baseko laginak *k* azpi-multzotan banatzen dira, parametroaren balioa 5 edo 10-en inguruan izan ei delarik. Azpi-multzoetako bat balioztatzeko, hau da, testeatzeko erabiltzen da, gainerako *k-1* azpi-multzoak entrenatzeko erabiltzen direlarik.

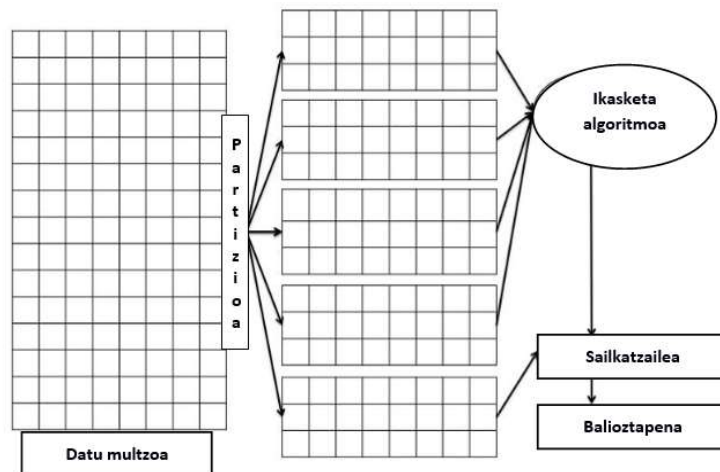
Orokorrean, datuen %90 entrenatzeko eta %10 balioztatzeko erabiltzen dira. Azal dutako datuen zatiketa beherago dagoen

Irudia 4k aurkezten du. Banaketa buruturik, balioztatze prozesua k aldiz errepikatzen da, balioztatze datu azpi-multzo posible ezberdin bakoitzarekin. Jarraian, balioztatze datuetarako iragarritako balioak balio errealekin egiaztatu eta iterazio bakoitzean lortutako emaitzen batez bestekoa kalkulatu da, emaitza bakarra izateko. Azkenik, lortutako eredua hautatutako metrika ezberdinekin ebaluatu egiten da.

Proiektu honen kasuan, gainera, k partiziodun balioztatze gurutzatua 100 aldiz errepikatu egiten da, ereduaren errendimendua ahalik eta eraginkorrena izatea lortu nahi baita.



Irudia 3. Algoritmo baten garapenean gerta daitezkeen hiru kasu ezberdinak. Lehenengo grafikoak underfitting-a argi aurkezten du; muga lerroak ez ditu grafikoko puntu guztiak aintzakotzat hartzen. Hirugarren grafikoak, kontrara, igarritako lerroak grafikoko puntu guztiak aintzat hartzen dituela aurkezten du; hasiera batean errendimendu oneko grafikoak dela pentsa daiteke, baina ez da horrela, zarata eta ohikoak ez diren puntuak ere kontuan hartzen ditu eta. Bigarren grafikoak, igarritako lerroa nahiko ona duen grafikoak da, puntuen gehiengoa kontuan hartzen dituelarik. Algoritmoen garapenean bilatu nahi den kasua, beti, azkeneko hau da [41].



Irudia 4. Datu multzoaren zatiketa Training eta Test azpi-multzoetan balioztatze gurutzatua erabilia [42].

5.4 Ebaluazio irizpideak

Ebaluazio neurriak ereduak aurretik ikusi ez dituzten laginak iragartzeko sailkatzaile eredu egokiena hautatzeko nahitaezkoak dira. Kasu honetan, nahasmendu matrizetik (ingelesez *Confusion*

Matrix, CM) kalkulagarriak diren sentikortasuna (ingelesez *Sensitivity, Sen*) eta igartze balio positiboa (ingelesez *Positive Predictive Value, PPV*) erabili dira algoritmo sailkatzaile ezberdinen errendimendua neurtzeko eta ebaluatzeko. Metrika hauek sailkapen bitarrentzako bakarrik baliagarriak direnez gero, [43] artikuluan aurkeztutako ikuspegia guk aztertuko dugun klase anitzen arazora orokortuak izan dira. Proiektua garatzeko erabilitako definizioak **Taula 1** kontingentzia taulan oinarritu dira. $N_{i,j}$ parametroak j klase moduan sailkaturiko (algoritmoaren emaitza) i klaseari dagokion lagin kopurua adierazten du (adituen erabakia). Kasu honetan, $i, j \in \{AS,ORG,TB,FB\} \equiv \{1,2,3,4\}$ definizioak erabili dira. Hortaz, i klaseko guztizko lagin kopurua R_i , eta j klasearentzako algoritmoaren guztizko erabaki kopurua C_j parametroak ezaguturik, hurrengo adierazpenak defini daitezke [2]:

$$R_i = \sum_{j=1}^4 N_{i,j} \quad \text{eta} \quad C_j = \sum_{i=1}^4 N_{i,j}$$

Gainera, guztizko lagin kopurua hurrengo adierazpenak lortuko du:

$$N_{tot} = \sum_{i=1}^4 \sum_{j=1}^4 N_{i,j} = \sum_{i=1}^4 R_i = \sum_{j=1}^4 C_j$$

Taula 1. Berpizte bihotz erritmoen sailkapenerako kontingentzia taula

		Algoritmoak iragarritako balioak				Σ
		AS	ORG	TB	FB	
Adituen erabakia	AS	$N_{1,1}$	$N_{1,2}$	$N_{1,3}$	$N_{1,4}$	R_1
	ORG	$N_{2,1}$	$N_{2,2}$	$N_{2,3}$	$N_{2,4}$	R_2
	TB	$N_{3,1}$	$N_{3,2}$	$N_{3,3}$	$N_{3,4}$	R_3
	FB	$N_{4,1}$	$N_{4,2}$	$N_{4,3}$	$N_{4,4}$	R_4
Σ		C_1	C_2	C_3	C_4	N_{tot}

Sailkatzaile bitarretan CM matrizeak 2x2 dimentsiodunak dira, eta honek definitzen dituen parametroak hurrengoak dira: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). TP parametroak positiboak diren eta ondo iragarritako lagin positiboak adierazten ditu, eta TN parametroak, kontrara, negatiboak diren eta ondo iragarritako lagin negatiboak adierazten ditu. FP parametroari dagokionez, lagin positibo moduan iragarritako, hau da, gaizki iragarritako lagin negatiboak dira, eta FN parametroak, lagin negatibo gisa iragarritako lagin positiboak adierazten ditu.

Proiektu honek ebazten duen klase anitzen problemarako, parametro hauek guztiak i klasearentzako jarraian aurkeztuko diren gisan orokortuak izan dira [2]:

$$TP_i = N_{i,i}$$

$$FP_i = C_i - N_{i,i}$$

$$FN_i = R_i - N_{i,i}$$

$$TN_i = N_{tot} - TP_i - FP_i - FN_i = N_{tot} - R_i - C_i + N_{i,i}$$

Horrez gain, sailkatzaileen errendimendua neurtzeko oso erabiliak diren Sen eta PPV metriken kalkulua i klase zehatz batentzako, beraz, hurrengo adierazpenek definitzen dute [2]:

$$Sen_i = \frac{TP_i}{R_i} = \frac{N_{i,i}}{R_i} \quad eta \quad PPV_i = \frac{TP_i}{C_i} = \frac{N_{i,i}}{C_i}$$

Hortaz, i klase zehatz batentzako sentikortasunak, algoritmoak berpizte erritmoaren sailkapen egokia egiteko probabilitatea adierazten du, hots, algoritmoak ondo sailkatzeko duen gaitasuna adierazten du. Iragartze balio positiboak, aldiz, algoritmoak igartzen duen berpizte erritmoa benetan pazienteak erritmo hori pairatzeko duen probabilitatea zein den adierazten du.

Orain arte aipaturiko parametroen definizioak ezaguturik, errendimendua neurtzeko eta CM matrizea parametro bakar batera murrizteko proiektuan erabilitako bi metrika orokorren definizioa egitea posiblea da: *Multiway Accuracy (MulAcc)* eta *Unweighted Mean of Sensitivities (UMS)* balioak [2]. Jarraian, hauetako bakoitzari egokitzen zaion ekuazioa aurkeztuko da hurrenez hurren:

$$MulAcc = \frac{1}{N_{tot}} \sum_{i=1}^4 N_{i,i} \quad eta \quad UMS = \frac{1}{4} \sum_{i=1}^4 Sen_i$$

Aurretik definitu direnez, MulAcc desorekarekiko oso sentikorra den, baina sailkatzaileak bere osotasunean aurkeztzen duen doiketa neurtzen duen parametroa da. UMSak, kontrara, sailkatu nahi diren erritmo klase guztiak pisu berdinarekin ebaluatzen ditu, klaseen arteko desorekak kalterik sor ez dezan.

Proiektuan garatzen diren sailkatzaile ezberdinen errendimendua neurtzeko azkeneko bi parametro hauek dira erabili direnak.

6. Alternatiben analisiak

Proiektuaren garapen eraginkorrerako eta helburuak modu eraginkorrean betetzeko existitzen diren alternatiba ezberdinak aztertzea da atal honen xedea. Lehendabizi, bihotz seinale desberdinen prozesaketa eta analisia, Machine Learning algoritmoen garapen egokia, baita hauek sortzeko beharrezko datu-base baten garapena ere ahalbidetzen duen programazio-lengoaia ezberdinetan oinarritutako software plataforma aukera desberdinak aztertuko dira, gero eraginkorrena aukeratuz. Ondoren, proiektuan proposatutakoa lortzea posiblea egiten duten ikasketa automatikoan funtsatutako algoritmo-sorta aztertuko da, hauen artean ere efizienteena hautatuz.

6.1. Programazio-lengoaia

Proiektuaren garapenerako software plataformari dagokionez, bost aukera gailenduko dira: Python, R, Matlab, Octave eta C programazio-lengoaiak hurrenez hurren. Hauetako bakoitzak aurkezten dituen ezaugarriak jarraian aipatuko dira, ondoren, irizpide ezberdinetan oinarrituta, proiektu honek proposatzen duen helburua lortzeko errendimendurik handiena aurkezten duen programazio-lengoaia eraginkorrena modu arrazoituenean aukeratzeko.

6.1.1. Python

Informatika zientzian Python programazio-lengoaia ezagunenetakoa eta erabilienetakoa da egun. Eskaintzen duen goi-mailako interaktibotasunak eta gero eta handiagoa den liburutegi zientifikoaren multzoak, algoritmoen garapenerako eta miaketa-datuen analisi sakonerako oso erakargarria bilakatzen dute [44].

Python sintaxi garbia, sinplea eta irakurterraza eskaintzen duen interpretatutako programazio-lengoaia da, C/C++ eta Java bezalako programazio-lengoaiekin alderaturik, konpilatzea beharrezkoa izan gabe, programaren exekuzioan zehar interpretatua baita. Objektuetara bideraturiko programazio inperatiboa zein programazio egituratua, plataformekiko independentea den scripting lengoaia, eta neurri txiki batean, programazio funtzionala jasaten duen paradigma anitzeko lengoaia da; hau da, lengoaia interpretatu eta multiplataformaduna zehazki. Gainera, testuinguru ezberdinetan modu doakoan erabiltzea posiblea egiten duen kode libreko lizentzia baten jabe ere bada.

Aurretik aipatu bezala, Python lengoaiak aurkezten duen abantaila nagusienetakoa bat zabala den liburutegi estandar batean eta interprete doakoan oinarritutakoa izatea da. Gai zehatzak aurrera eramateko programazio-lengoaiak berak liburutegi, datu mota eta funtzio kopuru handi bat barneratuta ditu, horrela, programen garapena arinduz. Proiektu honetan garatzen den Machine Learning-arekin loturik, programazio honek ikasketa automatikoan funtsaturiko ereduak entrenatzea ahalbidetzen duten liburutegi ugari ditu eskuragarri. Horrez gain, Python eta hau erabiltzeko beharrezkoak diren erreminta guztiak plataforma nagusietan eskuragarri daude, baita Python Web orrialde nagusian [45] ere, eskusibotasuna alde batera utziz. Hori dela eta, garapen-kostu altuak ordaindu nahi ez dituzten programatzaileentzako erakargarria oso den aukera bilakatzen da Python programazio-lengoaia.

Aldi berean, Python lengoaiak modulu eta paketeen erabilpena onartzen du, era horretan, programak modu modularrean diseinatuak izan daitezkeelarik eta kodea beste hainbat proiektuetan berrerabilgarriak izan daitezkeelarik ere; hots, modulu edo pakete bat sorturik, beste proiektu batzuetan erabiliak izateko eskalagarriak dira, eta modulu hauek inportatzeko eta esportatzeko zailtasunik ez daude [46].

Horrez gain, irakurgarritasunean oinarritzen den sintaxia duen ulerterraza eta ikasterraza den programazio-lengoaia sinplea da. Programatzaileei Python kodearen irakurketa eta itzulketara erraza egiten zaie beste lengoaia batzuekin konparaturik; honek, gainera, programen mantenu eta garapen kostuak asko murrizten ditu, lengoaia eta esperientzia mugarik gabe ekipoek elkarlanean egoteko gaitasuna baitute.

Dinamikoki idazten den programazio-lengoaia ere badela aipatzekoa da, horrek zenbait Python programatzailek diseinuan muga ugari ezartzea ekartzen duelarik, testatzeko beharrezko denbora luzeagoa izanik, eta erroreak, behin aplikazioak exekutatutako baino ez agertzen direlarik; hau da, erroreak exekuzioan besterik ezin dira hauteman [47].

Beste desabantaila batzuk ere aurkezten ditu: konpilazioa darabilten beste programazio-lengoaia batzuk baino geldoagoa izatea edota hauen produktibitatearen murrizketa besteak beste. Azken hau, beste programazio hizkuntzen aldean, zatikatuago egoteagatik gertatzen da.

6.1.2. R

R programazio-lengoaiak, Python programazio-lengoaiak bezala, datuen zientzia esparruan berebiziko garrantzia hartu du. Batez ere estatistika alorrean ezaguna den arren, ikasketa automatikoan oinarritutako gero eta kontzeptu eta funtzio gehiago barnean hartzen ditu, programazio honen erabilera eta garrantzia areagotuz. Funtzionaltasun zehatzak inplementatzen dituzten hainbat liburutegi ere existitzen dira, landu nahi diren programazio alderdi ezberdinak sakontzeko erraztasunak eskainiz.

Horrez gain, GNU lizentziadun software libre gisa banatutako multiplataformaduna den programazio-lengoaia ere bada; gainera, funtzionaltasun berrien etengabeko garapena dakarte, bertsio eguneratuagoak eta hobeak lortuz etengabe. Multiparadigma, objektuetara bideraturikoa eta bektoriala ere bada, Python moduan, programatzaile eta garatzaile erakunde handi batek erabiltzen duelarik, hau asko hobetuz eta aberastuz.

Esan bezala, R programazio-lengoaiak estatistika alorrean du jatorria, bektore eta matrizeekin eragiketarako egiteko gaitasun handia izanik, eta ondorioz, datu-baseekin lan egiteko erraztasun handia eskainiz; R-k, hortaz, datuak oso azkar manipulatzeko (aukeratu, birkodifikatu, berreskuratu) ahalmena du. Izatez, dagoeneko diseinaturik dauden datuak manipulatzera bideraturiko paketeak existitzen dira, aurrerago aztertutako eta analizatutako izango diren datuen prestaketan programazio-lengoaia iaioagoa eta eraginkorragoa izatea ahalbidetzen dutenak. Estatistika esparruan funtsaturiko programazioa izateagatik, aurkezten dituen beste abantailak batzuk datuen azterketarako agertzen dituen doitasuna eta zehaztasuna dira hurrenez hurren [48].

R programazioak, beraz, datuen manipulazio, grafiken bistaratze eta ikasketa automatikoan oinarritutako algoritmoen eskuragarritasunerako berezko ahalmena, batik bat, liburutegietan (sisteman bertan berez baneratutakoak batzuk eta instalatzeko aukerarekin beste batzuk) du.

Python-ekin alderaturik aurkezten duen desabantailetakoa bat, ikasketa kurba geldoagoa eta konplexuagoa izaten dela da.

6.1.3. Matlab

MATLAB (Matrix Laboratory) garapen integratuko ingurune (IDE, Integrated Development Environment) batean oinarritutako berezko programazio-lengoaia (M lengoaia) duen software matematikoa da. Gainera, ikerketa eta garapen erakundeetan, baita unibertsitateetan ere, erabilera oso zabalduta duen software multiplataformaduna da [49].

Izenak aditzera ematen duenez, MATLAB software plataforma, funtsezko osagai gisa matrizea duen sistema elkarreragilea da. Matrizeekin lan egiteko aukera honek, proiektu honi erraztasuna eskaintzen dio; izan ere, proiektu honen helburuetako bat, datu-base batean biltegituriko seinaleekin lan egitea eta datu-base baten garapena dira, aipaturiko ezaugarri hau oso lagungarria izanik. Horrez gain, interpretatutako lengoaia ere bada, horrela, berehalako eta momentuko “feedback”-a erabiltzaileari eskainiz, eta emaitzak edota sortutako erroreak momentuan erakutsiz; hau da, bistaratze eta arazketa erreminta eraginkorrak ditu.

Industria eremuan, MATLAB programazio-lengoaia ingeniartzan eta matematikan oinarritutako problema praktikoak ebazteko eta ikerketa esparruetan oso erabilia da, bereziki seinaleen prozesamendua eta kontrola darabilten aplikazioetan enfasi handia ezarriz. Horrez gain, programazio-lengoaia honek programatzaileentzako eta erabiltzaileentzako lagungarriak oso izan daitezkeen tresna osagarriak ere eskaintzen ditu. Horien artean, berriazko soluzioak hornitzen dituzten irudien prozesaketarako toolbox-ak, ordenagailu ikuspegia eskaintzen duten toolbox-ak, baita algoritmo-sorta zabala eskuragarri duten estatistika eta machine learning toolbox-ak ere daude. Aipaturiko toolbox hauek MATLAB erabiltzaileen gehiengoentzako garrantzia handia dute, bertan, problema jakinak ebazteko jadanik garaturiko funtzioak eta aplikazioak aurkitzen direlarik: hala nola, seinaleen prozesamendua, kontrol sistemen diseinua, sistema dinamikoen simulazioa, sistemen identifikazioa, neurona-sare artifizialak eta abar [50]. Bestalde, plataforma honek eskaintzen dituen funtzio ororen ezaugarri eta zehaztasunak eskuragarri dituen laguntza gida eskura dago, baita euskarri teknikoak hornitzen duen MATLAB web orrialde ofiziala ere [51].

Programazio-lengoaia honekin loturiko abantailekin jarraituz, mundu osoan oso zabalduta eta erabilia den programazioa izateagatik, proiektuarekin aurrera egin ahala ager daitezkeen arazoei aurre egiteko erraztasuna izatea posiblea da, seguruenik aurretiaz oztopo berdinak agertu eta hauentzako konponbide ugari web orrialdean aurki daitezkeelarik.

Software plataforma honetan aurki daitezkeen beste abantailetakoa bat, barnean hartzen duen eta programazioan asko laguntzen duen liburutegi-sorta da. Eskaintzen dituen liburutegien artean, liburutegi grafikoek berezko garrantzia dute, aztertutako datuen emaitzak kalitate oso altu batekin bistaratu, baita hauen aurkezpena erraztu ere egiten baitute. Proiektuan lortutako emaitzak plazaratzeko eta bistaratzeko, beraz, kalitate altuko emaitzak eskaintzen dituzten liburutegi grafiko hauek oso lagungarriak dira.

Proiektuaren garapenerako GUIaren (Graphical User Interface) erabilera egokia eta eraginkorra beharrezkoa da guztiz. Aurreko ataletan nabarmendu denez, proiektuaren helburu nagusia den sailkatzaile automatikoa burutzeko, aurretiaz, bihotz seinale datu-base zabal baten analisi sakona eta zehatza efizienteki gauzatea ezinbestekoa da; horretarako, MATLAB plataformak, GUIDE deritzon interfaze grafikoak eraikitze ingurune eraginkorra, aurreratua eta erabilterraza eskaintzen du, horrela, GUIaren kalitatea ahalik eta onena izatea lortuz.

Software plataforma honek, ordea, abantailak bakarrik ez, desabantailak ere agertzen ditu. Alde batetik, interpretatutako lengoia izateagatik, konpilazioa darabilten lengoaiak baino geldoak izatea gertatzen da, abiadura ezberdintasuna oso altua ez den arren. Hala ere, desabantailarik nagusiena erosteak duen prezio altua da. Toolbox tresna kontuan hartu gabe, erabiltzaile bakarrerako MATLAB lizentzia batek 2000€ inguru balio ditu, erosten den toolbox bakoitzeko 1000€ inguru ordaindu behar direlarik [52]. Normalean proiektuetan hiruzpalau toolbox nahitaezkoak izaten direla kontuan harturik, MATLAB software plataformaren prezioa 5000€ bitartekoa izatera ailegatu daiteke.

6.1.4. Octave

Octave edo GNU Octave kalkulu numerikoak jorratzeko programa edo programazio-lengoia, MATLAB software plataformaren baliokide den kode libreko doako software-a da. Octave eta MATLAB programazioek elkarrekin partekatzen dituzten ezaugarrien artean, biek interprete bat eskaintzen dutela gailentzekoa da, horrela, modu interaktiboan aginduak exekutatu ditzaketelarik [53].

Plataforma biek eskaintzen dituzten aukerak, erremintak eta funtzioak oso antzekoak diren arren, ezberdintasun nabariak ere agertzen dituzte. Octave programazioak eskaintzen duen interfaze grafikoa, adibidez, ez da MATLAB programazioak eskaintzen duen interfaze grafikoa bezain eraginkorra, intuitiboa eta erabilerraza zoritxarrez. Ildo honetatik, Octave plataformak interfazearen garapenerako eskaintzen dituen tresnak mugatuak dira, proiektu honetan garatu beharreko GUIak duen garrantzia kontuan harturik.

Horrez gain, software plataforma honek eskuragarri dituen funtzio eta tresna osagarriak, ez dira MATLAB plataformak eskaintzen dituenak bezain lagungarriak; gainera, plataforma honek eskuragarri dituen toolbox kopurua, MATLAB plataformak dituenak baino urriagoak dira, azkeneko programazio-lengoia honetan sortu behar ez diren eta beharrezkoak izan daitezkeen funtzioak Octave plataforman programatzea nahitaezkoa bilakatuz.

Desabantailei amaiera emanez, mundu mailan Octave MATLAB programazio-lengoia bezain hedatua ez dagoela, eta ondorioz, euskarri tekniko oso egokia ez dela aipatzekoa da. Ez hori bakarrik, honek eskaintzen duen laguntza gida MATLABek eskuragarri duena bezain osatua, zehatza eta erabilerraza ez dela ere aipatzekoa da. Arrazoi hauek direla medio, erabiltzaileei suertatzen zaizkien erroreak eta arazoak konpontzeko zailtasun handiagoak izatea oso posiblea da.

Laburbilduz, MATLABek eskuragarri dituen euskarri eta baliabideak Octave programazioak eskain ditzakeenak baino ahaltsuagoak, aberatsagoak eta eraginkorragoak dira, konplexutasun maila altuko proiektuetan laguntza eta erraztasun handiagoa hornituz.

6.1.5. C

C programazio-lengoaia Sistema eragileen inplementaziora bideraturiko lengoaia da. Honako programazio-lengoaia ezaguna da, batez ere, lengoaia konpilatua izateagatik sortzen duen kodearen eraginkortasuna eta exekuzio azkarra dela medio, baita software sistemak garatzeko programazio lengoaiarik erabiliena izateagatik ere; horrez gain, aplikazioak gauzatzeko ere oso erabilia da. Hortaz, aurretik eztabaidatu izan diren interpretatuak diren programazio-lengoaiekin alderaturik, C programazioa efizienteagoa dela ondoriozta daiteke. Horrez gain, irudiekin, grafikoekin zein datu-baseekin lan egiteko erraztasuna eskaintzen duten matrizeen manipulazioa ez dago eskuragarri, MATLAB edo Octave programazio-lengoaiekin gertatzen ez den bezala. Definituriko funtzioak barnean hartzen dituen liburutegiak ere eskuragarri daude C programazioan; hauek, ordea, aurreko lengoaiek eskuragarri dituzten toolbox-ak bezain eraginkorrak eta aurreratuak ez dira. Horregatik, tresna berezien garapena nahitaezkoa bilakatzea posiblea da, lanaren jorraketa motelduz, baita zailtasun maila ere handituz.

Programazio mota honek aurkezten duen malgutasuna, aipatzekoa den beste abantailetakoa bat da; honek, era ezberdinetan eta plataforma ezberdinetan programatzea ahalbidetzen du. Horrez gain, kodea exekutatzeako software plataformarik behar ez duenez, dohainik programatzeko aukera ere badago.

Lengoaia bisuala ez dela aipatzekoa da ere. C programazio-lengoaia ez da intuitiboa, eta honek barnean hartzen dituen arauak argi egon behar dute, programazioan zailtasunak eta oztopoak aurki ez daitezken; hots, C programazioari ahalik eta etekinik handiena atera ahal izateko, honen erabilera xehetasun handiarekin ezagutzeak esanguratsuki laguntzen du. Programazio honen ikasketa kurba, beraz, geldoa dela esan genezake [54][55][56].

6.1.6. Software plataformaren hautaketa irizpidea

Atal honetan, deskribaturiko software plataforma guztien artean proiektuaren xede nagusia arrakastatsu lortzeko eraginkorrena eta errendimendurik handiena aurkezten duena hautatzeko, aintzakotzat hartu diren irizpide ezberdinak adieraziko dira, bakoitzak proiektuarekiko duten garrantzia ere deskribatuz. Kasu honetan, proiektuaren garapen egokirako programazio-lengoaiek, bakoitzak bere aldetik, duten eragina hurrengo irizpideekin ebaluatu da: erraztasuna, datuen antolaketa eta bistaratzea, algoritmoaren garapena, eraginkortasuna eta prezioa.

➤ **Erraztasuna**

Software plataformaren konplexutasuna da irizpide honekin ebaluatzen dena, garatzen den proiektuarekiko duen eragina kontuan harturik eta garrantzia neurtuz.

Pisua: 2/10

➤ **Datuen antolaketa eta bistaratzea**

Proiektu honetan datu kopuru handia biltegitratzen duen datu-base batekin lan egiten da; hori dela medio, datu-basearekin lan egiteko erraztasuna eta laguntza eskaintzen duen software plataforma

hautatzea oso garrantzitsua da, proiektuaren garapen eraginkor baterako datuen analisi sakon eta egoki bat burutzeko ere. Aldi berean, bistaratzen diren grafiken kalitatea altua izatea nahitaezkoa da, lortzen diren emaitzen azterketa sakona gauzatu eta ondorio baliotsuak lortzeko.

Pisua: 3/10

➤ **Algoritmoaren garapena**

Ikasketa automatikoan oinarritutako algoritmoen garapenaren erraztasuna eta eraginkortasuna ebaluatzeak ere garrantzia du. Horretarako, machine learning-ean funtsatutako algoritmoek barnean hartzen dituzten liburutegien aniztasuna eta boterea kontuan hartu dira.

Pisua: 2/10

➤ **Eraginkortasuna**

Software plataformak kodea exekutatzekoan aurkezten duen efizientzia aintzat hartzekoa da ere.

Pisua: 2/10

➤ **Prezioa**

Prezioa beti kontuan hartu beharreko faktorea da; hala ere, proiektu honen garapen arrakastaturako beste faktore batzuek garrantzi handiagoa jaso dute.

Pisua: 1/10

Jarraian, gorago deskribaturiko programazio-lengoaia bakoitzak aipaturiko irizpide bakoitzeko duen pisua laburbiltzen duen taula aurkezten da, ondoren, proiektuaren garapenerako efizienteena hautatzeko:

Taula 2. Software plataformaren aukeraketa egiteko irizpide ezberdinen arabera ebaluaketa laburbiltzen duen taula

	Irizpidea	Pisua	Python	R	Matlab	Octave	C
	<i>Erraztasuna</i>	2/10	1.5	1.5	2	1.5	1
	<i>Datuen antolaketa eta bistaratzea</i>	3/10	2	2	3	2.5	1
	<i>Algoritmoaren garapena</i>	2/10	2	1.5	2	1.5	0.5
	<i>Eraginkortasuna</i>	2/10	1	1.5	1.5	1.5	2
	<i>Prezioa</i>	1/10	1	1	0.1	1	1
	Guztira	10/10	7.5/10	7.5/10	8.6/10	8/10	5.5/10

Taula 2ko emaitzak aztertuz, Matlab eta Octave software plataformak oso antzekoak direla agerikoa da; ezberdintasunik esanguratsua Matlab plataformaren prezioa litzateke, oso altua dela jakina dena. Proiektu honek helburu duen sailkatzaile automatikoaren garapenerako, ordea, prezioa ez den gainerako faktoreek eragin handiagoa dute: datu-baseak bildurik dituen datuen analisi eraginkorrerako interfaze grafikoak eraikitzeak erraztasuna izatea, datuak antolatzeko eta maneiatzeko gaitasuna erdiesteak, eta programak sortzeko eskaintzen dituen abantailak izatea besteak beste. Aipaturiko honengatik guztiarengatik, Matlab izan da proiektuaren garapenerako aukeratutako software plataforma.

6.1. Sailkatzaile automatikoak

Ikasketa automatikoak eskuragarri dituen sailkatzaile automatikoak garatzeko algoritmoak, sailkatu nahi diren sarrera datuak deskribatzen dituzten ezaugarri esanguratsu eta adierazgarrietan oinarritzen dira, ikasketa mota hau ikasketa gainbegiratu gisa ezaguna delarik; hau da, aurretik ezagunak eta kategorizatuak izan diren datu multzo batean oinarrituta algoritmoa entrenatuz eta ebaluatuz honen garapena lortu egiten da [57]. Machine learning-ak eskaintzen dituen sailkatzaile automatiko ezberdinen artean eraginkorrena hautatzeko, analizatu nahi den datu-multzoa ezagutu, sailkapenerako erabiliko diren datuak ezaugarritzen dituzten elementuak ezagutu, baita datu kopurua eta ezaugarrien dimentsioak ezagutzea ere ezinbestekoa da. Proiektuaren helburua jakinik, hurrengo bost algoritmoak nagusitu dira: K-nearest neighbors (K-NN), Erregresio logistikoa (EL), Support Vector Machines (SVM), Artificial Neuronal Networks (ANN) eta Random Forest (RF).

6.1.1. K-nearest neighbors (K-NN)

K-NN algoritmoa elementuen espazioan hurbileko adibideen entrenamenduan oinarritutako gainbegiratutako sailkapen metodoa da.

Sailkapen sistema honek duen oinarritzko funtzionamendua antzekotzat jotzen diren elementuen klasea esleitz elementu berriak sailkatzean funtsatzen da. Horretarako, algoritmoaren entrenamendu fasean zehar aztertutako k objektu hurbilenean klasea frogatu egiten da, analizatzen den elementua aipaturiko k objektuen klaseen artean gehien errepikatzen den motan sailkatuz.

Algoritmo honen exekuzioa existitzen diren sailkatzaile algoritmoen artean sinpleenatarikoa den arren, praktikan oso ahaltsuak suertatzen dira, emaitza onak lortuz eta erabilienatarikoa ere izanik.

Algoritmo honek, beraz, aurkezten dituen abantailen artean, exekutatzeko sinplea eta,aldi berean, eraginkorra dela aipatzekoa da. Ez hori bakarrik, datuen banaketarekiko suposiziorik ere egin behar ez dela, eta entrenamendu fasea bizkorra dela nabarmentzekoak dira.

Desabantailak ere, ordea, egon badaude. Alde batetik, ez du eredurik sortzen, algoritmoa sortzeko garrantzia duten eta kontuan hartzen diren ezaugarriak klase ezberdinekin duten harremana ulertzeko abilezia mugatuz. Bestetik, proiektuak duen helburuaren arabera k balio egoki bat aukeratzea beharrezkoa da. Gainera, sailkapen fasea geldoa da oso, eta ezaugarri nominalek, baita galdutako datuek ere, prozesamendu gehigarria behar izaten dute, prozesua are geldoagoa izatea eraginez [58].

6.1.2. Erregresio logistikoa (EL)

Erregresio logistikoa (ingelesez Logistic Regression, LR) medikuntza alorreko auresanetan oso erabilia den sailkapenerako erabiltzen den eredu linealetan oinarritzen da. Aldagai independenteen arabera aldagai kategoriko (kategoria kopuru mugatua har dezakeen aldagaia hain zuzen ere) baten emaitza aurreikusteko baliatzen den metodo mota da. Beste faktore batzuk medio jazotako gertaera baten probabilitatearen eredu egiteko oso erabilgarria izaten da [58].

Lehendabizi, beraz, eredu lineala klase edo kategoria zehatz bat gertatzeko probabilitatera doitzen da, ondoren, klase edo kategoria zehatz hauetako baten emaitza zehazten den atalase bat sortzeko

funtzio bat erabiltzeko. Erabaki muga kurba logaritmiko edo esponentzial batera aldatu daiteke, horrela, doikuntza hobea lortuz [59].

Sarrera gisa linealki banangarriak izan behar diren datu erregularrak beharrezkoak ditu; hori dela eta, analizatuko diren datuak antolatu eta atipikoak ez diren balioak ezabatu behar dira, metodo honek overfitting-a pairatzeko joera izaten baitu. Horregatik, datuen aurretiko bistaratzea eta analisisa derrigorrezkoa da guztiz.

Ezaugarri berrietara moldatzeko metodo honek duen erraztasuna aurkezten duen abantailetakoa bat da, baita datu kopuru handi batekin lan egiten denean oso azkarra dela ere.

6.1.3. Support Vector Machines (SVM)

Erregresio eta sailkapen problemekin loturiko gainbegiraketa ikasketan oinarritutako algoritmoen multzoa da SVM. SVM batek dimentsionaltasun altuko espazio batean hiperplano bat zein hiperplano multzo bat eraikitzen du; hau da, SVM algoritmo batek, analizatzen diren datuen talde ezberdinak banatzen dituen plano bat bilatzea du xede. Banaketa honek, planoaren aukeraketak planoan dauden puntu hurbilenen arteko aldea maximizatzen duela aditzera ematen du; aipaturiko puntu hauek, bektore euskarri deritze [59].

Eredu mota honek aurkezten dituen abantailen artean, “zaratatsuak” diren datuek eragin askorik sortzen ez dutela, eta hainbat aldiz aipaturiko overfitting arazoa izateko joera izaten ez duela da. Beste algoritmo batzuk baino erabilerrazagoak direla ere aipatzekoa da, sare neuronalak baino errazagoak kasuan kasu. Gainera, gero eta erabiliagoa da mundu mailan, alde batetik, eskaintzen duen zehaztasunagatik, baita datuen analisisira bideraturiko ikerketetan izaten duen arrakasta dela medio ere.

Ahultasunak ere baditu; eredurik egokiena eraikitzeak, parametro ezberdinen arteko konbinaketa desberdinen hainbat froga egitea ekartzen du. Sarrerako datuek adibide edo parametro asko izatekotan, gainera, entrenamendua geldoa izan daiteke. Hori gutxi balitz, ematen dituen emaitzak interpretatzeko konplexuak, edota ezinezkoak ere izatera ailegatu daitezke [58].

SVM metodoa testu sailkapen problemekin ezaugarri sailkapen problemekin baino eraginkorragoa dela aipatzekoa da. Dimentsio handiko espazioak norma diren problemetan bereziki, zehaztasun eta doitasun altua aurkezten du.

6.1.4. Artificial Neuronal Networks (ANN)

Artificial Neuronal Networks (ANN) algoritmo eredu, gizakiaren garunaren portaerari eta egiturari buruz matematikan formalizazioak erdiesteko saiakera gisa sortu zen eredu konputazionala da. Esperientzian oinarritutako ikasketa da, horretan funtsaturik behar beste ezagutza erauziz eta eskuratuz.

ANN algoritmo batek duen berriarazko xede nagusia, gizakiak sor ditzakeen errore edo akatsak ekiditea posiblea egiten duen prozesatzaile artifizialen bitartez sistema zentral biologiko baten emulazioa da. Sistema biologiko neuronal baten oinarritzko osagaiak neuronak dira, milaka neuronez osatutako sareetan taldekatuta eta geruzetan oinarritutako egitura batean antolatuta daudenak. Sistema neuronal artifizial batean antzeko sistema hierarkiko bat ezartzea posiblea da, modu horretan, elkarren artean loturik dauden zein kanpo-sarreretara loturik dauden eta seinalea hainbat bidetatik

hedatzea posiblea egiten duen irteera bat duten oinarrizko prozesatzaileen (neurona artifizialak) bilduma batean bilakatuz [60].

Sarearen loturak aleatorioki hasieratzen dira, eskuragarri dauden datuekin entrenatzen diren heinean modu mailakatu batean doituz, eta modu horretan, entrenamendurako erabilitako datu multzoaren kasu guztiak arian-arian ezagutzera eta aurrez aurre ikasiz [61].

Sare neuronal batek hartzen duen egitura eta dituen oinarriak direla eta, burmuinarekiko oso antzekoak diren ezaugarri kopuru handi bat aurkezten du: lehen aipatu legez, esperientziatik ikasteko, aurreko kasuak egoera berrietara orokortzeko eta garrantzirik gabeko informazioa eskaintzen duten sarreretatik abiatuz funtsezko ezaugarriak lortzeko gaitasuna aurkezten du besteak beste. Honek, abantaila ugari eskaintzea eragiten du, baita esparru eta alor ezberdinetara aplikagarria den teknologia izatea ere. Eskuragarri dituen abantailen artean, hurrengoak aipagarrienak dira [62]: sendotasuna eta akatsekiko tolerantzia, zehaztugabea zein lausotua, zaratatsua, osatugabea, hutsala eta funtsik gabeko informazioa maneiatzeko eta erabiltzeko aukera, orokortzeko gaitasun altua, ikasketa egokitua, denbora errelean exekutagarria izatea eta dagoeneko existitzen den egungo teknologian egokitzeko erraztasuna aurkezten ditu. Gainera, entrenatzeko datu-base zabala behar dute, eskalagarriak dira eta konfiguratzeko errazak ere badira. Datuen prozesaketarako, ordea, denbora gehiago behar dute.

Medikuntza esparruan ANN algoritmoek aplikazio anitz dituzte, baita seinaleen prozesatzean oinarritzen diren telekomunikazioen esparruan ere. Eskain ditzakeen aplikazio guztien artean, irudien antzematean, ahotsaren ezagutzan eta hizkuntza-prozesamenduan dira batez ere erabiliak, soluziorik eraginkorrenak lortuz.

6.1.5. Random Forest (RF)

Aurretik deskribatu denez, Random Forest teknika sailkatzaile independente bakoitzaren eraikuntzan aleatoriotasuna gehituz sailkapenean doitasuna eta zehaztasuna hobetzen duen Leo Breiman-ek garaturiko agregazio teknika da. Aipaturiko ausazkotasuna, zuhaitzaren garapenean zein entrenamendu laginean txertatu daiteke. Random Forest algoritmoa, beraz, datu multzo batetik abiatuz erabaki zuhaitz batzuetan oinarrituta eraikitzen den metodoa da, amaierako iragarpena zuhaitzen botoen gehiengoak erabakitzen duelarik.

Problema mota ezberdinetan aplikagarria den algoritmoa da, zaratatsuak diren datuekin lan egiteko ahalmena du, eta bereizgarriak eta benetan garrantzitsuak diren ezaugarriak hautatzeko gaitasuna aurkezten du. Gainera, datu-base zabalekin eta datuak ezaugarritzen dituzten parametro kopuru handi batekin modu efiziente batean funtzionatzeko ahalmena du. Ausazko basoen exekuzioa bizkorra da oso, eskalagarria da guztiz, eta zehaztasuna eta doitasuna mantenduz desorekatutako eta desagertutako datuak tratatzeko aukera ere aurkezten du. Aipatutako abantaila hauek guztiek, RF ikasketa automatikoa algoritmo trebeenetariko izatea eragiten dute. Horrez gain, desabantaila nagusia, teknika honek sortzen duen eredia interpretatzeko nahiko konplexua dela da [63].

6.1.6. Sailkatzaile automatikoaren hautaketa irizpidea

Atal honetan, deskribaturiko sailkatzaile automatiko guztien artean proiektuaren xede nagusia arrakastatsu lortzeko eraginkorrena eta errendimendurik handiena aurkezten duena hautatzeko, aintzakotzat hartu diren irizpide ezberdinak adieraziko dira, bakoitzak proiektuarekiko duten garrantzia ere deskribatuz. Kasu honetan, proiektuaren garapen egokirako algoritmo sailkatzaile bakoitzak bere aldetik duen eragina hurrengo irizpideekin ebaluatu da: ereduaren malgutasuna, algoritmoaren abiadura, interpretatzeko erraztasuna, doitasuna eta zehaztasuna.

➤ *Ereduaren malgutasuna*

Sailkatzailearen garapenerako nahitaezkoak diren parametro kopurua, eta hauen konplexutasun maila da ezaugarri honekin neurtuko dena. Honekin loturik, algoritmoaren erraztasuna ere kontuan hartuko da.

Pisua: 3/10

➤ *Algoritmoaren abiadura*

Algoritmoak entrenamendu faserako, baita ebaluazio faserako ere, beharrezkoa duen denbora ebaluatu nahi da. Horrez gain, eredia behin garaturik, hau erabiltzeak ekartzen duen iraupena ere kontuan hartzekoa da.

Pisua: 3/10

➤ *Interpretatzeko erraztasuna*

Algoritmoa garaturik hau interpretatzeko zailtasun maila ebaluatzea ere garrantzitsua da.

Pisua: 2/10

➤ *Doitasuna eta zehaztasuna*

Ezaugarri honen bitartez, algoritmoak eskaintzen duen sailkapenerako doitasun eta zehaztasun mailak neurtuko dira.

Pisua: 2/10

Jarraian, lehenago deskribaturiko ikasketa automatikoan oinarrituriko sailkatzaile bakoitzak aipaturiko irizpide bakoitzeko duen pisua laburbiltzen duen taula aurkezten da, ondoren, proiektuaren garapenerako efizienteena modu arrazoituan hautatzeko:

Taula 3. Sailkatzaile automatikoaren aukeraketa egiteko irizpide ezberdinen arabera ebaluaketa laburbiltzen duen taula

<i>Irizpidea</i>	<i>Pisua</i>	<i>K-NN</i>	<i>EL</i>	<i>SVM</i>	<i>ANN</i>	<i>RF</i>
<i>Ereduaren malgutasuna</i>	3/10	2.5	1.5	2	2	3
<i>Algoritmoaren abiadura</i>	3/10	1.5	2.5	1.5	2	2.8
<i>Interpretatzeko erraztasuna</i>	2/10	1	1.5	0.8	1.5	1
<i>Doitasuna eta zehaztasuna</i>	2/10	1	1	1.8	2	2
<i>Guztira</i>	10/10	6/10	6.5/10	6.1/10	7.5/10	8.8/10

Taula 3-ko emaitzak aztertutik, proiektu honetarako erabiliko den algoritmorik egokiena Random Forest ikasketa automatikoa dela agerikoa da. Aukera honek emaitzak interpretatzerakoan zailtasunak aurkezten dituen arren, proiektu honen xedea erdiesteko beste faktore batzuek garrantzi eta eragin handiagoa dutela aipatzekoa da, bereziki, algoritmoak aurkezten duen malgutasuna, erraztasuna, abiadura eta doitasuna gailenduz.

7. Arriskuen analisia

Proiektu baten arriskuen kudeaketak gertaera positiboen probabilitatea eta eragina handitzea, eta gertaera negatiboen probabilitatea eta eragina txikitzea du helburu. Proiektuaren amaieran arrakasta erdiesteko, proiektua garatu ahala arriskuen kudeaketa eta analisia modu proaktibo, trinko eta sendo batean jorratzea ezinbestekoa bilakatzen da. Arriskuak proiektua sortu bezain laster existitzen dira, eta arriskuen kudeaketari modu proaktibo eta aktibo batean hasiera eman gabe proiektua ekiteak, proiektuan gerta litezkeen arriskuen eragina eta probabilitatea haztea ekar lezake, horrek, proiektuaren porrota probableagoa izatea lortuz [64].

Atal honen xedea, beraz, proiektu honen garapenean zehar zein proiektu honen garapenak izan ditzakeen arriskuen identifikazioan, aurreikustean, analisisian, kudeaketan eta kontrolean datza; hau da, ager daitezkeen mehatxu posibleen eta desiragarriak ez diren gertaeren kausen azterketan, baita hauek ekar ditzaketen kalteen eta ondorioen analisisian ere, oinarritzen da. Horrela, arriskuak izan ditzaketen hedapena, elkarrekiko lotura eta garrantzia ahalik eta xehetasun handienaz ikertzea lortu nahi da.

Arriskuaren definizioak berak oinarrizko bi aldagai aztertzea eramaten zaitu: alde batetik, arriskuak berez gertatzeko duen probabilitatea, eta beste alde batetik, arrisku horrek gure proiektuan izan dezakeen eragina; hots, arriskuak sor ditzakeen galerak. Bideragarritasun azterlana den arriskuen analisi hau aurrera eramateko, beraz, lehendabizi analisi kualitatibo bat gauzatuko da, ondoren, ustekabeen aurrean erantzun plan bat burutuz; hau da, kontingentzia plan bat garatuz. Jarraian, aipaturiko probabilitate eta inpaktu hauek neurtu ahal izateko kontuan hartuko diren irizpideak laburbiltzen dituzten taulak eskuragarri daude:

Taula 4. Probabilitatearen neurketarako irizpide taula deskribapena barne

Irizpidea	Pisua	Deskribapena
Baxua	0,2	Arriskua gertatzea ekar dezakeen edota probabilitate ertaina zein altua erdietsi dezakeen arrazoi pisutsurik ez da existitzen
Ertaina	0,5	Hau gertatzeko maiztasuna larria ez izateko oinarriak existitzen dira
Altua	0,8	Arriskuak maiztasun handiarekin gertatzeko aukera du

Taula 5. Inpaktuaren neurketarako irizpide taula deskribapena barne

Irizpidea	Pisua	Deskribapena
Baxua	0,1	Arrisku bat existitzen dela uste bada ere, proiektuaren helburuarekiko eta proiektuarekiko kalte txikia eragingo luke, neurriak hartzea beharrezkoa izan gabe
Ertaina	0,5	Analisiak arriskua garrantzitsua dela esaten badu ere, honek ez du helburua ezta honekin loturiko emaitzak guztiz kaltetuko
Altua	0,9	Arriskua gertatzekotan, helburuarekiko kalte handia eragingo luke

Jarraian, proiektuan identifikatutako eragina izan ditzaketen arriskuak eta gertaerak bana-banan aipatuko dira gertatzeko probabilitate maila ere aipatuz:

A. Kostuen arriskuak

Kostuen arriskuen multzoaren barnean ager daitezkeen arrisku ezberdinen artean, gure proiektuaren kasuan, proiektua bera garatu ahala aurreikusi gabeko gastuek proposatutako garapen kostuak gainditzea, baita hasieratze fasean balioetsitako kostuak egokiak ez izatea ere gerta liteke. Aipaturiko arrisku hauek, prezioen aldaketak egoteagatik, aurreikusi gabeko lan gehigarriak jorratu behar izateagatik, materialen aldaketak zein material gehiago beharrezkoak izateagatik eta abar ager daitezke.

Aipaturiko gertaera hauek guztiak ekiditeko, eta beraz, kostuen estimazio arrazoitsua lortu ahal izateko, proiektuan zehar gauzatzen den atazetako bakoitza sakonki ebaluatu da. Horrez gain, aurrekontua aldatzea saihestea posiblea ez bada, aurrekontuan ezustekoak deritzon kontu saila gehitzen da, kasu honetan %5ekoa izanik. Hala ere, proiektu honen garapenerako espresuki materialik erosi ez dela aipatzekoa da, proiektu osoaren jorraketa software plataforma batean oinarrituta baitago.

Arrazoi hauek guztiak direla medio, arrisku hau suertatzeko probabilitatea eta proiektuarekiko eragina, biak, oso txikiak dira:

Probabilitatea: Baxua Inpaktua : Baxua

B. Egutegi arriskuak

Proiektuan zehar proposatutako lan planean zehaztutako datak zein iraupenak gainditzea, zenbait ataza garatu eta ebazteko arazo teknikoak zein kanpo arazoak izatea edota esfortzu handiagoa behar izatea, baliabideen esleipena egokia ez izatea zein aurretik antolatu eta kontuan hartu ez diren baliabideen esleipena izatea, proiektu honetan aurreikus daitezkeen egutegi arriskuen egoera ezberdinak dira. Lan plana aurretik planifikaturik egon arren, eginkizunak landu ahala espero ez diren aipatutako kanpoko arazoak zein arazo teknikoak suertatu daitezke, hauen ondorioz, alde batera utzi ezin diren atzerapenak ekarriz.

Arrisku hauen gertaera probabilitatea altua da oso, hauek gertatzekotan ekiditea ezinezkoa izanik. Hori dela eta, arrisku hauen eragina eta probabilitatea murrizteko, lan-fluxua proposatzean ataza bakoitzari esleitu beharreko iraupena egokia izateko xeheki aztertu da. Horrez gain, ataza bakoitzaren kudeaketa aproposa lortzeko helburuarekin, proiektuaren antolaketarako arduraduna den lan-pakete bat ere definitu da. Kasu honetan, proiektuko lantaldeak bi pertsonak besterik osatzen ez dutenez, zeregin baten atzerapenak proiektua osoa arriskuan jartzea oso konplexua da, atzerapen horrek partaide baten lana bakarrik atzeratuko bailuke, galdutako ordu horiek oso erraz berreskuratzea posiblea izanik.

Probabilitatea: Altua Inpaktua : Baxua

C. Arrisku teknikoak

Proiektuan garatu beharreko software plataformaren edota datu-basearen kalitateak zein aldi baterako planifikazioak mehatxuak izatea eta kudeaketa plataformaren diseinuan, inplementazioan, interfazean, baieztapenean edota mantentzean arazoak identifikatzea dira arrisku teknikoen artean aurki daitezkeen gertakizunetako batzuk. Horrez gain, plataforma eta datu-basea sorturik, algoritmoaren garapeneko kalitate zein planifikazio arazoak ere arrisku teknikoen barnean sartzen dira.

Proiektuak xede duen berpizte bihotz seinaleen sailkatzaile automatikoaren garapenerako, lehendabizi lortu beharreko analisi kudeaketa software plataformak ezinbestekoa den kalitatea ez betetzekotan zein diseinuan edota interfazean arazoak, akatsak identifikatzekotan, proiektuaren jomugara ailegatzeko gauzatu beharreko hurrengo pausuetan eragina izango luke; hau da, ondorengo pausuetan jorratu beharreko datu-basearen, baita algoritmoaren garapenaren kalitatean ere eragingo luke, proiektu osoaren eta emaitzen kalitate maila asko murriztuz edota proiektuaren helburua ez lortzera ere iritsiz. Proiektuak jasango lukeen inpaktua, beraz, altua izango litzateke, honen probabilitatea ertaina izanik.

Proiektuan lanean ari den taldea txikia izateagatik, komunikazioa handiagoa da, eta proiektuaren garapen zuzenerako atazetan esfortzu handiagoa nabarmentzea errazagoa da. Hala ere, software kudeaketa tresnaren diseinurako programazio kodea oso luzea da, eta baliteke programa erabiltzaileentzako nahi bezain dinamikoa ez izatea erritmoen analisien prozesua motelduz.

Probabilitatea: Ertaina
Inpaktua : Altua

D. Arrisku teknologikoak

Proiektua garatzeko beharrezkoak diren teknologien erabilpenean ezagutzak ez izatea edo teknologia berrien konplexutasunen ulermenean zailtasunak izatea, proiekturako garatu nahi den funtzionalitaterako egokiak ez diren erreminten erabilpena, hau da, alternatiben aukeraketa okerra, hardware/software arazoak izatea eta proiektuan paraleloan garatutako interfaze edo script fitxategien integrazioan arazoak izatea besteak beste, arrisku teknologikoak diren multzoaren barnean hartzen dira.

Kasu honetan, analisira bideraturiko software tresna sortzeko erabili beharreko Matlab GUI interfazearen inguruko prestakuntza bat egiten da; horrela, programazio ingurune honetan funtsezkoak diren kontzeptuak bereganatzeko, ezagutzeko eta erabiltzeko erraztasuna eskaintzen da. Nahiz eta GUI interfazearen erabilpena lehendabizikoa ez izan, hasiera batean erabiltzeko zailtasunak aurkitzea posiblea da oso; dena den, hasiera batean programazioan ordu gehiago emanaz arazoa erraz konponduko litzateke. Aurreko atalean aipaturiko alternatiben hautaketa okerra egitekotan, erdietsi nahi diren helburuak lortzeko oztopoak eta

zailtasunak izatea, edo zuzenean, helburuak ez lortzea gerta liteke. Aipaturiko azkeneko kasu honetan, alternatibaren analisia berriro egin, aukeraketa zuzena egin, eta berriro ere, proiektuari hasiera eman beharko litzaioke. Gainera, proiektuan erabilitako ekipoek funtzionatzeari uzteak edota tresnen aukeraketa okerra egiteak, aurreikusitako gabeko gastuen igoera eta baita bete beharreko eginkizunen atzerapena ekarriko luke, proiektuarekiko eragina handia izanik. Honakoa saihesteko, alde batetik, alternatibaren azterketa xehea egiteak berebiziko garrantzia dauka, eta bestetik, ekipoak behar bezala erabiltzearekin eta mantentzearekin probabilitatea asko murriztuko litzateke. Azkenik, interfaze bakarra erabiltzeagatik ez luke integrazio arazorik izan behar, script fitxategi bat baino gehiago erabili arren. Hortaz, esandako guztiarekin, arrisku teknologikoak gauzatzeko probabilitatea baxua dela eta inpaktua altua dela esango genuke.

Probabilitatea: Ertaina
Inpaktua : Altua

E. Operazio arriskuak

Operazio arriskuen barnean hartzen diren egoerak hurrengoak dira: proiektua garatu ahala ager daitezkeen arazoei egokia ez den ebazpidea ematea, proiektuan lan egiten duen taldearen lidergo eskasia egotea, komunikazio gutxi egotea, taldeak izan behar duen motibazioa txikia izatea, monitorizazio, inplementazio eta garapen arazoak egotea eta abar.

Proiektu honetan zentratuz, honen garapenean arazoak ekartzen dituzten gertaerak agertzekotan eta hauen ebazpidea okerra izatekotan, argi dago proiektuarekiko eragina altua litzatekeela; hala ere, arazo hauek era egoki eta lasaiago batean kudeatzeko aukera dago, batez ere lan taldea oso txikia izateagatik, partaideak berak denbora nahi bezala kudeatu baitezake. Ildo honetatik, hau gauzatzeko probabilitatea baxua da. Lidergo arazoa, komunikazio arazoa zein taldearen motibazioa txikia izatearen arazoa egoteko probabilitatea ere txikia da, askotan aipatu legez, lan taldea bi pertsonak bakarrik osatzen dutelako. Gainera, ezohikoa den arren, proiektuan lanean dagoen partaideren baten baja egotea gerta liteke, honek proiektuarekiko edukiko lukeen inpaktua oso altua izanik; izan ere, lan taldea bi pertsonak besterik ez dute osatzen eta proiektuaren garapenean zeharreko edozein ataza, arazo eta abar tirabiran jarriko luke. Horrez gain, bajaran dagoen langile horren eginkizun guztiak ezingo luke beste batek gauzatu, alde batetik, atzerapenak sortuko bailirateke, eta bestetik, baja hartu duen partaide horren erantzukizun eta betebeharrak beste partaide baten ardurapean jartzea posiblea ez baita; honek, beraz, proiektuarekiko inpaktuan eragin altua izango luke. Arrisku honi aurre egiteko, proiektuan parte-hartzen duten langileen ezagutza maila duten, baita proiektuaren garapenerako nahitaezkoa den erantzukizunen arduradun izateko gai diren ordezkari langile adituak eskuragarri izatea nahitaezkoa da. Monitorizazio, inplementazio eta garapen arazoei dagokienez, software plataformaren, datu-basearen zein algoritmoaren garapen, monitorizazio edota inplementazio kaskarra egiteak, argi dago proiektua bere osotasunean gaitzetsiko lukeela, sekulako atzerapenak ere eraginez. Horregatik, proiektua osatzen duten ataza ezberdinak xehetasun eta zehaztasun osoz burutu behar dira.

Probabilitatea: Ertaina
Inpaktua : Altua

F. Merkatu eta kompetentzia arriskuak

Gerta liteke kardiologia eremuan oinarritutako, eta batez ere, berpizte bihotz seinaleen sailkatzaile automatikoetan oinarritutako gaur egungo egoerako merkatuak aldaketak izatea, eta hori dela medio, proiektuan garatutako algoritmoa zein metodoa zaharkituta gelditzea eta erabiltzerik ez izatea. Horrez gain, esparru honetan oinarritutako mundu mailako kompetentzia altua dela aipatzekoa da; zientzialari eta ikerkuntza erakunde asko dira bihotz-biriketako geldialdien aurreko egoeren arreta eta berpiztea ahalik eta kalitate eta eraginkortasun handienekoa izateko berpizte bihotz seinaleen sailkatzaile automatikoetan ikertzen ari direnak. Algoritmoak ugariak dira, inplementazio eta kudeaketa modu ezberdinak aplikagarriak direlarik. Argi dago, beraz, aipatutako arrisku hauek guztiek proiektuan izango luketen inpaktua oso handia izango litzatekeela, eta arazo hauek gertatzeko maiztasuna larria ez izateko oinarriak dituela.

Probabilitatea: Ertaina
Inpaktua : Altua

G. Lege aldaketak

Lege aldaketak, araudi aldaketak edota proiektuan eragina izan ditzaketen estandarrak agertzea ere posiblea litzateke, baina gaur egun, medikuntza alor honi loturiko araudi, lege zein estandarrak zehatzak dira eta finkaturik daude, aldaketak izateko oso probabilitate txikia izanik. Aipaturiko azkeneko kanpoko arrisku honi dagokionez, proiektuan izango lukeen inpaktua oso handia izango litzatekeela agerikoa da.

Probabilitatea:Baxua
Inpaktua : Altua

Behean, aipatutako arrisku hauen probabilitate-eragina erlazioa aurkezten duen matrizea eskuragarri dago.

Matrize honi esker, arriskuek beraien aldetik proiektuan emateko duten probabilitatea eta arrisku bakoitza jasoz gero proiektuan zer nolako eragin maila duten aztertzea oso era erraz batean egitea ahalbidetzen du.

Argi dago, arriskuen kontrolak, edozein proiektu batean atzerapenak edota porrotak ekar ditzakeen faktore guztiak kontuan hartzeagatik, edozein proiektu baten arrakasta bermatzen duela. Proiektu baten garapenean zeharreko planifikazioan arrisku kontrol egoki bat egitea oso garrantzitsua da, beraz, honek, proiektua ezarritako kalitate eta iraupenetan garatzera bultzatzen duelako.

		ERAGINA		
		Baxua 0,1	Ertaina 0,5	Altua 0,9
PROBABILITATEA	Baxua 0,2	A Baxua (0,02)	Baxua (0,1)	G Ertaina (0,18)
	Ertaina 0,5	Baxua (0,05)	Ertaina (0,25)	F, E, D, C Altua (0,45)
	Altua 0,8	B Baxua (0,08)	Altua (0,4)	Altua (0,72)

Irudia 5. Proiektuan eragina duten arriskuen analisiari egokitzen zaion probabilitate-eragina matrizea.

8. Soluzioaren deskribapena

Atal honetan, proiektuaren helburu nagusia den berpizte bihotz erritmoak identifikatzen eta sailkatzen dituen algoritmo automatikoa garatzeko jarraitutako fase ezberdinak eta metodologia sakon azalduko dira. Honakoa erdiesteko, hiru azpi-helburu gauzatzea nahitaezkoa izan da: lehendabizi, berpizte bihotz seinaleen analisisian oinarritutako anotazioa eta segmentazioa egitea beharrezkoa izan da, horrela, algoritmoaren garapena bultzatzeko ezinbestekoak diren datuen bilduma lortzeko. Bigarren azpi-helburua, aurreko fasean analizatutako eta aztertutako seinale segmentu egokiak bakarrik biltzen dituen datu-base baten sorrera izan da; honek, lortu nahi den ikasketa automatikoan oinarritutako algoritmoaren garapenerako beharrezko informazioa eta datuak biltegiturik ditu. Azkeneko fasea, proiektuaren xede den ikasketa automatikoan funtsaturiko algoritmo eraginkor baten garapena da, horretarako, aurretik, ebaluaketa eta hautaketa pautak ezberdinak ere gauzatuz.

Lehenago aipatu denez, proiektuaren garapenerako Matlab software plataformak eskuragarri duen programazio ingurune bisuala erabili da, hau da, erabiltzailearen interfaze grafikoan oinarritu da. Jarraian, beraz, interfaze bakarrean garatutako eta jarraitutako hiru faseen metodologia deskribatuko da:

8.1. 1-FASEA: BERPIZTE BIHOTZ SEINALEEN ANALISIAN OINARRITUTAKO ANOTAZIOA ETA SEGMENTAZIOA

Sistema kardiobaskularrarekin eta arnas sistemarekin loturiko edozein terapia bat tratatzeko, berpizte gidetik gomendatzen dituzten teknika eta metodo ezberdinen eraginkortasuna, kalitatea eta fidagarritasuna areagotzeko, baita hauek hobera egitea bilatzen eta ahalbidetzen duten ikerketak errazteko eta sustatzeko ere, egoera horien aurrean pazienteek pairatzen dituzten bihotz seinaleen eboluzioa biltzen duten datu-baseen analisi sakonak berebiziko garrantzia du. Gainera, edozein Machine Learning algoritmo baten garapenean erabiliak eta aztertuak izango diren datuen bilketa eta azterlana, burutu beharreko lehenengo pausua da.

Aipatutako datu-baseko erregistro bakoitzak jasotzen duen informazioa egituretan antolatutarik dago, paziente bakoitzari dagokion erregistroak bihotz-biriketako geldialdia jasandako gaixoen datuak jasotzen dituelarik. Jarraian, *physio_ann.mat* deitutako datu-baseko egitura bakoitzak duen informazioa aurkezten duen taula eskuragarri dago:

Taula 6. *Pazienteen EKG seinaleen physio_ann.mat datu-base fitxategiaren erregistroen egitura*

Izena	Datu mota	Azalpena
reg_name	string	Paziente bakoitzari dagokion erregistroaren izena
ann	array<cell>	KDA gailuak pazientearen EKG seinalean adierazitako eta grabaturiko mezua
t_ann	array<double>	KDA gailuak adierazitako mezuen t segundo aldiuneak
t_comp	array<double>	Bihotz-biriketako berpizte konpresio bakoitza burutu den t segundo aldiuneak

Taula 6an ikus daitekeenez, KDA gailuek pazienteen seinaleak grabatzeaz gain, hauek automatikoki momentu oro eta egindako analisi bakoitzeko egoerari buruzko informazioa ematen duten mezuak, hauen denbora posizioak, baita BBB maniobren denbora posizioak ere eskaintzen dituzte, ondoren, seinaleen azterketa eta analisisa egiteko erraztasunak emanez.

Datu-base honetaz gain, *resumenDatosExt.mat* fitxategiak pazienteen bihotz seinaleen inguruko informazio gehigarria metaturik duen eta datu hauen kudeaketa errazten duen taula bat eskuragarri dago. Taula honek, KDA tresnek egindako seinale bakoitzeko analisi seinale tartek berezi, eta horietako bakoitzari dagokion informazioa eskaintzen du, bakoitza bere aldetik. Azkeneko honen egitura hurrengoa da beraz:

Taula 7. *Pazienteen seinaleen inguruan informazio gehigarria ematen duen resumenDatosExt.mat fitxategiaren egitura*

Izena	Datu mota	Azalpena
Registro	string	Analisi tarte bakoitzari dagokion pazientearen erregistroaren izena
Fecha	double	Analisisa egindako data
Analisis	double	Paziente bati egokitzen zaion seinaleak barnean hartzen dituen seinale bakoitzaren analisi zenbakia
DiagAED	double	KDA gailuak seinalearen azterketaren ondoren, pazienteari deskarga elektrikoa emateari buruz harturiko erabakia
RitmoCons	uint8	Medikuek adostutako (consenso) erritmo sailkapena
AnalisisOK	double	KDA gailuak harturiko erabakia zuzena den ala ez adierazten du
Comentarios	string	Medikuek anotazio prozesuan egindako iruzkinak
RitmoMedicos	array<uint8>	Alorrean adituak diren 6 mediku ezberdinen erritmo diagnosis
CommentsMedicos	array<cell>	Seinaleen analisisan medikuei suertaturiko egoerari buruzko informazioa ematen dituzten banakako iruzkinak

KDA desfibrilagailuen bitartez seinaleak eta hauen egoeraren inguruko informazioa bildurik, bihotzarekin loturiko medikuntza esparruan adituak diren mediku ezberdinen seinaleen azterlana gomendatzen da, hobekuntzak eta aurrerakuntzak lortzeko nahitaezkoak diren ondorengo analisisetan oso lagungarriak dira eta. Horregatik, proiekturako erabilitako azkeneko taula honek hornitzen dituen datu esanguratsuen artean, adituak diren sei medikuen analisiak eta anotazioak aurki daitezke, baita azkenean adostutako erritmo sailkapena ere.

Era berean, KDA gailuetatik lorturiko proiektuaren garapenerako ezinbestekoak diren paziente bakoitzeko seinaleez eta hauei buruzko informazioaz osaturiko *.mat* fitxategiak ere eskuragarri izan ditugu, hauetako bakoitzaren egitura hurrengoa izanik:

Taula 8. *KDA gailutik eskuraturiko paziente bakoitzeko seinaleen informazioa gordetzen duen .mat fitxategi bakoitzaren egitura.*

Izena	Datu mota	Azalpena
channels	struct	EKG seinalea eta inpedantzia seinaleari dagokion kanala hurrenez hurren
signals	struct	KDA gailutik eskuraturiko seinaleen anplitudeak, baita hauen laginketa frekuentzia ere

Aldi berean, paziente bakoitzeko .mat fitxategietako bakoitzak, barnean bi egitura ere badituela Taula 8an ikus daiteke. Kasu honetan, aztertuko diren pazienteen egitura hauek guztiek dituzten barneko aldagaiak berdinak dira:

Taula 9. *Paziente bakoitzeko KDA gailuak eskuraturiko datuen artean **channels** egiturak barnean hartzen dituen datuen balioak*

Izena	Datu mota	Azalpena
imp	double	Inpedantzia seinalea lortzeko erabilitako KDA gailuaren kanala
ecg	double	EKG seinalea lortzeko erabilitako KDA gailuaren kanala

Taula 10. *Paziente bakoitzeko KDA gailuak eskuraturiko datuen artean **signals** egiturak barnean hartzen dituen datuen balioak*

Izena	Datu mota	Azalpena
s_imp	array<double>	Inpedantzia seinalearen anplitude balioak
fs_imp	double	TTI inpedantzia seinalea lortzeko erabilitako laginketa maiztasuna
s_ecg	array<double>	EKG seinalearen anplitude balioak
fs_ecg	double	EKG seinalea lortzeko erabilitako laginketa maiztasuna

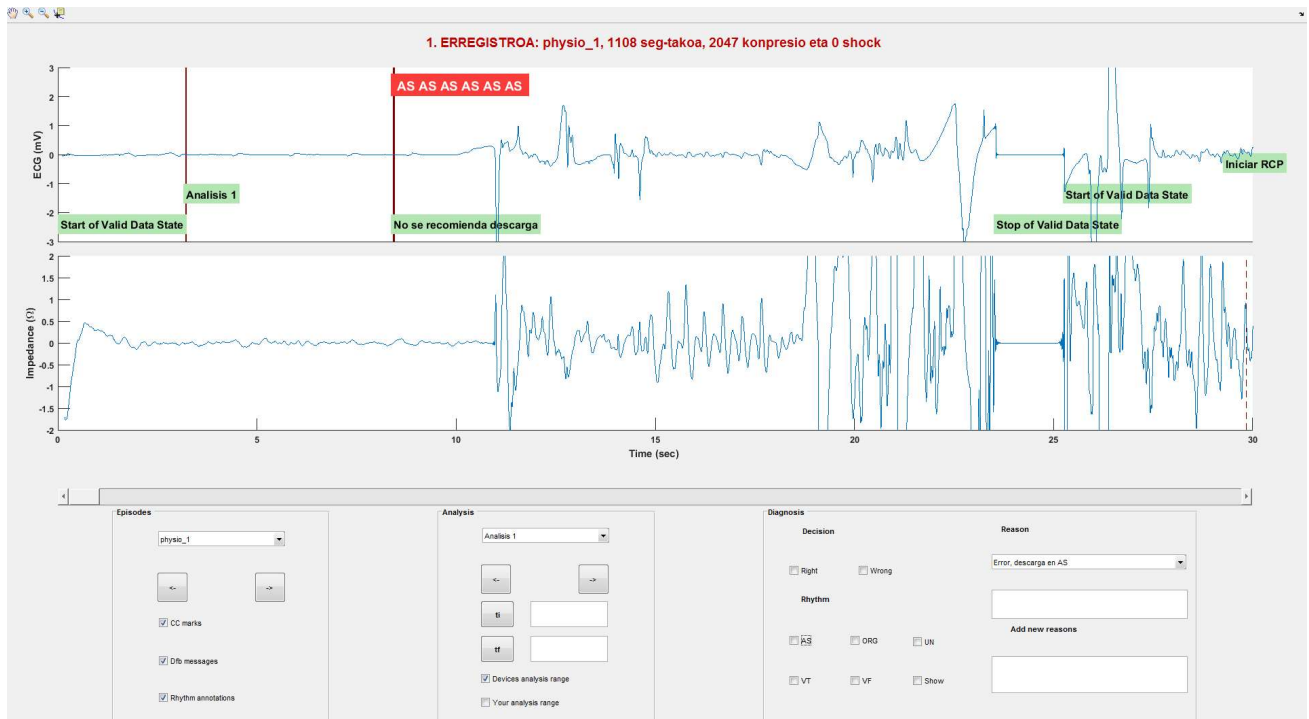
Proiektuari hasiera emateko eskuragarri dauden datuak aurkezturik, berriro ere lehenengo fase honi hel diezaiokegu. Aipaturiko Osakidetza erakundeak utzitako ospitaletik kanpoko bihotz-biriketako geldialdia jasan duten 857 pazienteen erregistroen bistaratzea, azterketa eta anotazioa egiteko aukera izateko tresna baten garapena nahitaezkoa dela agerikoa da, ondoren, sailkatzaile automatikoaren osaketarako ezinbestekoa den erregistro horien anotazioa eta segmentazioa ahalbidetzeko. Lehenengo fase hau, beraz, bereizitako bi azpi-atal ezberdinetan deskribatuko da: seinaleen kudeaketarako tresna lortzeko jarraitutako pausuak eta hauen metodologia, baita honek eskaintzen dituen funtzionalitate ezberdinak ere, deskribatuko den lehendabiziko atala izango da, jarraian, datu-basea osatzen duten seinaleen banakako analisia, azterketa eta segmentazioa egiteko irizpideak eta metodologia ere aipatuz.

8.1.1. Kudeaketa tresnaren garapena

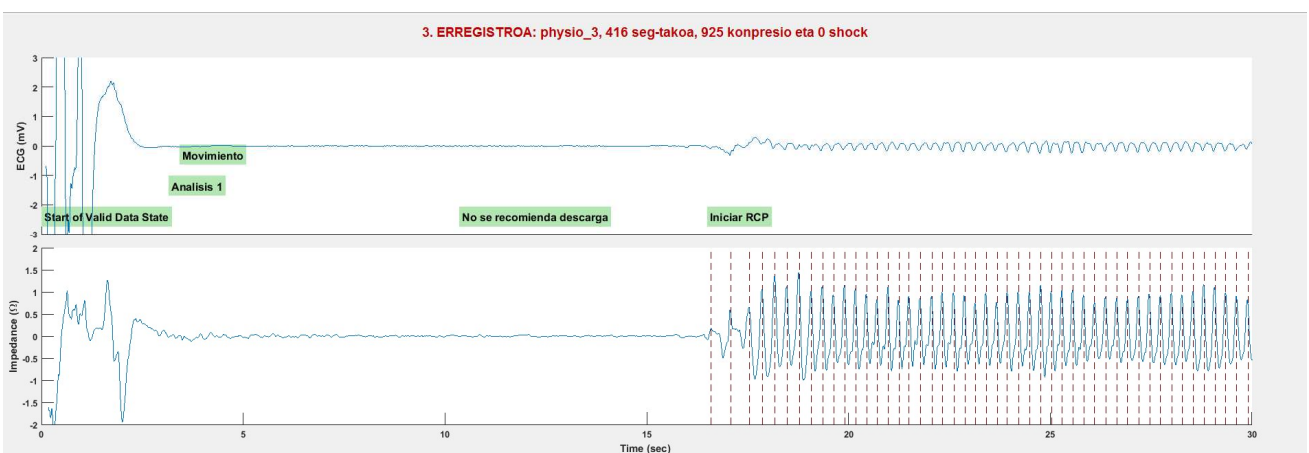
Behean eskuragarri dagoen Irudia 6ak, proiektuaren garapenerako ezinbestekoa den eraikitako kudeaketa tresna bere osotasunean, hots, plataformaren egitura eta eskuragarri dituen seinaleen kudeaketarako funtzionalitate ezberdinak aurkezten ditu. Eraikitako software kudeaketa tresna honen helburu nagusia, berpizte bihotz erritmoak sailkatzeko ikasketa automatikoan oinarritutako algoritmo baten garapenerako hasieran ezinbestekoa eta beharrezkoa suertatzen den datu-base efiziente bat sortzeko bihotz-biriketako geldialdia jasan duten gaixoen seinaleen bilakaeren analisi eta azterketa eraginkor, aberats eta sakona egiteko, erraztasunak eta laguntza eskaintzea da.

Tresna honek, beste tresna batzuekiko erakusten duen ezberdintasun eta abantailetakoa bat, pazienteen EKG seinalea bistaratzeaz gain, informazio gehigarria eta lagungarria ematen duen seinale horri dagokion TTI (Transthoracic Impedance) inpedantzia ere erakusten duela da. Seinale honek BBB bihotz-informazio handia ematen du; alde batetik, KDA momentuan ondo konektaturik dagoen ala ez

jakitea ahalbidetzen du, eta bestetik, laginketa maiztasuna handia denean ($f_s > 50$ Hz) pazientearen momentu bakoitzean bular-sakadak jasotzen dituen ala ez ere jakinarazten du, interferentzia hori inpedantzia nabarmentzen baita. Proiektu honetan lantzen den datu-basea KDAek ematen duten informazioan oinarriturik egoteagatik, alde batetik, gaixo bakoitzari dagokion EKG seinalea eta TTI seinalea bistaratuko dira, EKG seinalean KDAak momentu oro adierazitako mezuak eta BBB maniobrak noiz gauzatu diren ere agerraraziz; bestetik, TTI inpedantzia seinalean konpresioak ere aurkeztuko dira. Informazio honekin, proiektuan emaitza hobekiak lortzeko helburuarekin, bihotz-biriketako geldialdien egoeren ondorengo analisi aberatsago bat egitea ahalbidetzen da. Jarraian, azaldutako hau erakusten duen paziente zehatz bati dagokion adibidea Irudia 7an eskuragarri dago.

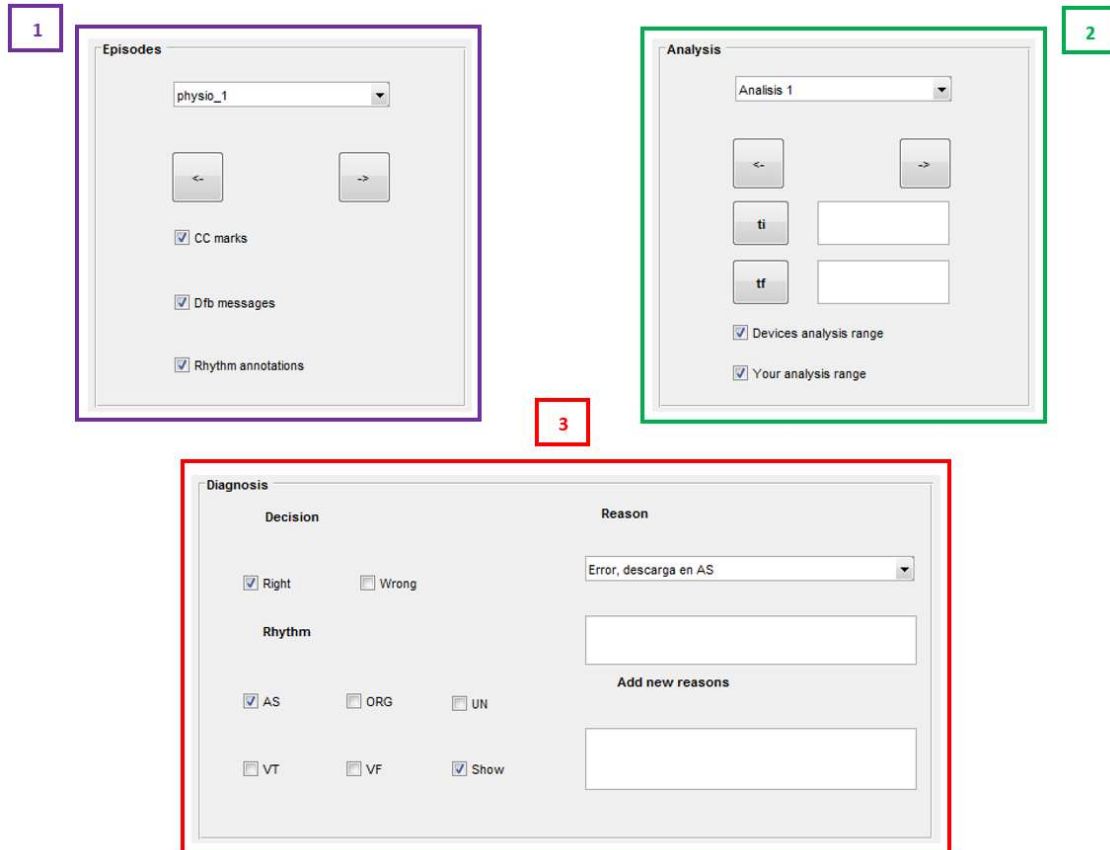


Irudia 6. Berpizte bihotz seinaleen bistaratzea, analisia eta kudeaketa egiteko gaitasuna eskaintzen duen Matlab-en funtsaturiko software plataforma tresna



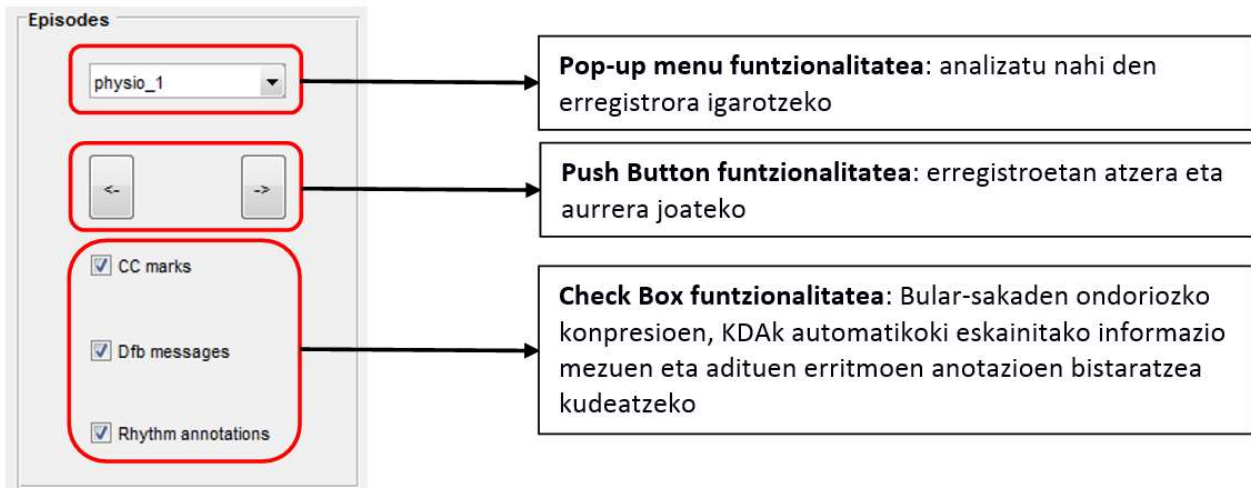
Irudia 7. Datu-baseko hirugarren erregistroari dagokion pazientearen EKG eta TTI seinaleak, baita KDA gailuaren egoera mezuak eta BBB maniobraren ondoriozko konpresioak ere agertzen dira. Goiko grafikoan pazienteari dagokion EKG seinalearen lehenengo 30 segundoen bilakaera ikusten da; bertan, KDAak seinalearen azterketan zehar automatikoki aurkeztzen dituen mezuak ere agertzen dira, bakoitza egokitzen zaion aldiunean. Beheko grafikoak, seinale tarte horri berorri dagokion inpedantzia aurkeztzen du, pazienteari emandako bular-sakadekin bat

Seinaleen, mezuen eta konpresioen bistaratzeaz gain, datuen kudeaketarako funtzionalitate ezberdinak ere eskaintzen ditu; zehazki, beheko irudian nabarmenduriko gertaeretara, analisietara eta diagnosira bideraturiko hiru azpi-panel ezberdinen funtzionalitate desberdinak aurkezten ditu:



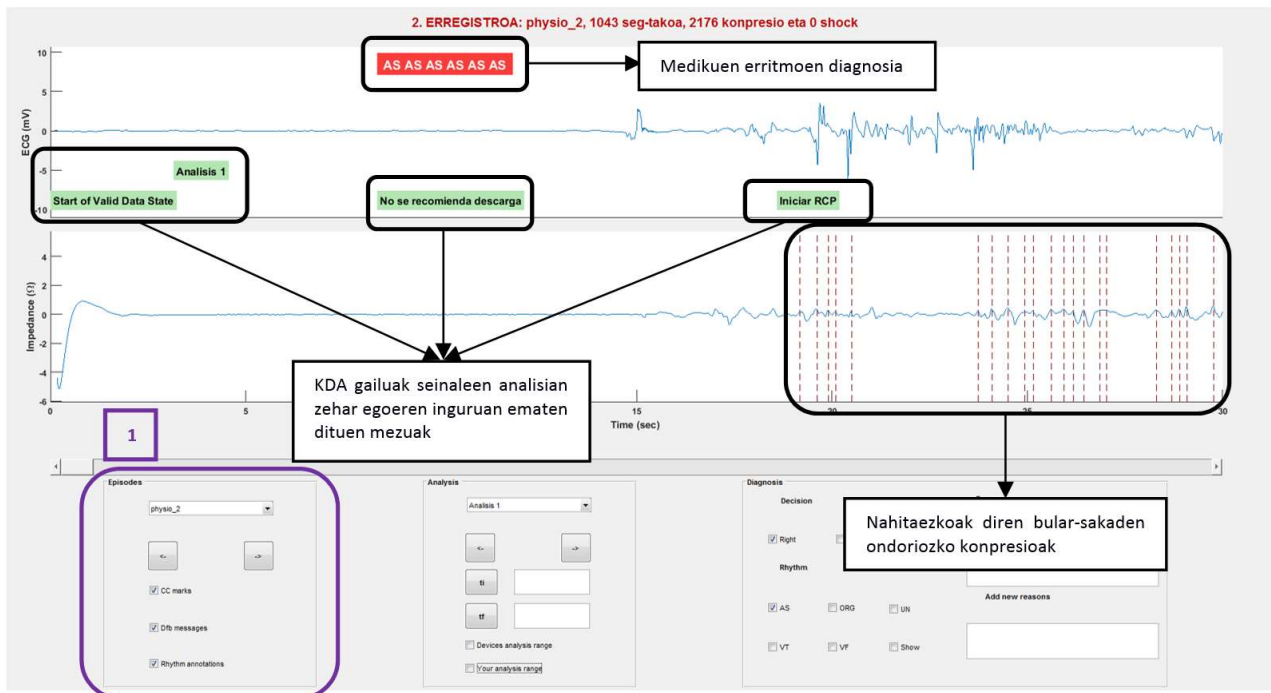
Irudia 8. Garaturiko kudeaketa tresnak gertaeretarako, analisietarako eta diagnosirako eskaintzen dituen funtzionalitate ezberdinak

Datu-baseak biltegiturik dituen 857 paziente ezberdinen artean bistaratu nahi den erregistroa hautatzeko aukera ematen duen *Pop-up menu* funtzionalitatea da gertaeretara bideraturiko lehenengo azpi-panelak eskuragarri duen funtzionalitateetako bat. Horrez gain, erregistro batetik bestera aurrera eta atzera joateko erraztasuna ematen duten bi *Push Button* ere erabilgarri ditu. Aurretik aipatu denez, proiektuaren garapenerako eskuragarri izan dugun datu-baseak, KDA gailuak pazienteen seinaleen bilakaeran zeharreko egoera ezberdinei buruz eskaintzen dituen mezuak, baita bular-sakaden konpresioen denbora uneak ere batzen ditu; horrez gain, KDA gailuak buruturiko analisien seinale tarte berdinen azterketa eta diagnosia egin duten sei mediku ezberdinen emaitzak ere erabilgarri daude; era horretan, adituek pazientearen bihotz seinale osoan zehar adierazitako eta detektaturiko erritmo usteak erraz ezagutu eta aztertu daitezke. Ildo honetatik, aipaturiko azkeneko hiru ezaugarri hauen bistaratzea kudeatzeko aukera ere badagoela nabarmentzekoa da, horretarako, GUIak eskaintzen duen *Check Box* funtzionalitatea erabili delarik. Jarraian, gertaeretara bideraturiko funtzionalitate ezberdinak agertzen dituen azpi-panela bistaratu da:



Irudia 9. Gertaeretara bideraturiko lehenengo azpi-panelak eskuragarri dituen funtzionalitate ezberdinen laburpena

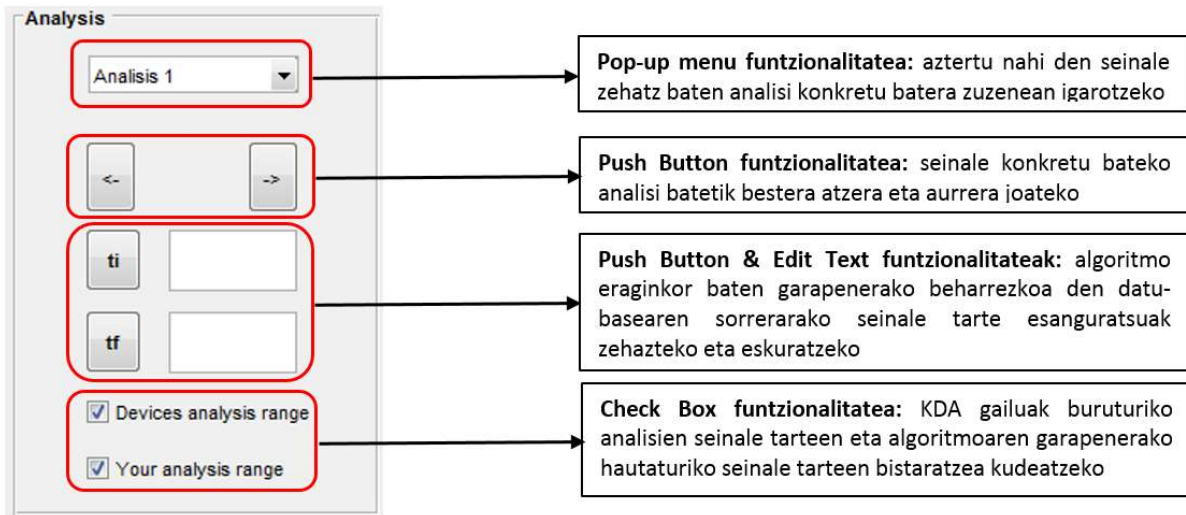
Behean ageri den Irudia 10ean, orain arte deskribaturiko gertaeretara bideraturiko funtzionalitate aktibo ezberdinetan oinarriturik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren analisiaren emaitzak ageri dira. Alde batetik, aipaturiko KDA gailuaren seinalearen lehenengo analisisan funtsaturiko mezu automatikoen arabera, bihotz seinale tarte horrek erritmo hilgarririk pairatu ez, eta gaixoak deskarga elektrikorik behar izan ez duela ikus daiteke. Era berean, 20.segundoan inguru, BBB maniobrei hasiera eman, eta bular-sakaden ondorioz inpedantzia seinalean suertatzen diren interferentziak ere ageri dira. Azkenik, proiektu hau aurrera eramateko oinarri harturiko 6 mediku adituen erritmoen diagnostia ezagutzeko aukera ere badago; kasu honetan, aditu guztiek bat etorri, eta bihotz seinale horren lehenengo analisi horri dagokion tarte asistolia erritmoari dagokiola, eta KDA gailuaren azkeneko erabakiarekin bat datozela ere, nabarmentzekoa da.



Irudia 10. Gertaeretara bideraturiko lehenengo azpi-paneleko funtzionalitateak aktibo egonik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren azterlanaren emaitza

Badakigunez, datu-basean biltegituriko paziente bakoitzak pairaturiko bihotz seinaleen bilakaeran zehar burutzen den azterlana hainbat analisi ezberdinetan antolatuz dagoela. Ildo honetatik, analisisetara bideraturiko bigarren azpi-panelak, erregistro bakoitzak jorraturik dituen seinale tarte desberdinen analisisen artean, aztertu nahi den seinale tarte konkretu bateko analisi batera zuzenean jotzeko aukera ematen duen *Pop-up Menu* funtzionalitatea eskuragarri du alde batetik. Horrez gain, eta aurretik deskribaturiko lehenengo azpi-panelak bezalaxe, honek ere analisi batetik bestera atzera eta aurrera joateko aukera ematen duten bi *Push Button* ere erabilgarri ditu.

Software kudeaketa plataforma honek, berpizte bihotz seinale baten bilakaeran zehar KDA gailuak gauzatzen dituen analisi ezberdinen denbora posizio tarteak ere ezagutzeko, maneiatzeko eta kudeatzeko aukera ematen du. Alde batetik, gailuak berak analizia noiz hasi eta noiz bukatu duen ezagutzeko zehazten duen seinale tarte bataratzeko aukera eskaintzen duen *“Devices analysis range”* izendatutako *Check Box* funtzionalitatea eskuragarri du. Bestetik, aurrerago azalduko den eta proiektuaren garapenean funtsezkoa eta ezinbestekoa den seinale hauen guztien azterketa eraginkor eta aberats bat gauzatzeko xedearekin, guk geuk algoritmoaren jorretarako esanguratsua eta egokiak ikusten ditugun seinale tarteak hautatzeko gaitasuna eskaintzen duten zenbait funtzionalitate eskuragarri daude. Honakoa lortzeko, *“ti”* eta *“tf”* deituriko bi *Push Button*, baita bi *Edit Text* elementu ere, eskuragarri ditu. GUI plataformak eskaintzen duen *Push Button* funtzionalitatean oinarrituriko *“ti”* eta *“tf”* elementuak sakatzuz, eta horren ondorioz, EKG seinalean agertzen den kurtsore batez baliaturik, algoritmoaren garapenerako erabiliko den datu-baserako seinale tarte garbi, argi, zuzen eta esanguratsua zehaztea ahalbidetzen da; zehatzago, *“ti”* sakatzuz eskuratu nahi den bihotz seinale tartearen hasierako unea eta *“tf”* sakatzuz eskuratu nahi den bihotz seinale tartearen amaierako unea adierazten dira hain zuzen ere. Aldi berean, *“ti”* eta *“tf”* *Push Button* elementuen bitartez zehazten diren hasiera eta amaierako unek, alboan aurki daitezkeen *Edit Text* elementuetan agertzen dira ere. Horrez gain, *Edit Text* elementu horietaz baliatuz, eskuratu nahi den seinale tarte nabarmenen hasierako eta amaierako denbora unek eskuz zehaztea ere posiblea da. Analisisetara bideraturiko azpi-panel honek eskaintzen duen azkeneko funtzionalitatea, KDA gailuak buruturiko seinale analisi tarteen bataratzea kudeatzeko aukera dagoen bezala, erritmoen sailkatzaile automatikoaren garapenerako erabakitako seinale tarteen bataratzea kudeatzeko aukera ere eskaintzen du; azkeneko hau ere, GUIak eskuragarri duen *Check Box* elementua erabili da, kasu honetan, *“Your analysis range”* deiturikoa. Jarraian, aipaturiko azkeneko azpi-panel honek eskaintzen dituen funtzionalitateen deskribapena laburturik aurkezten dituen irudia eskuragarri dago:



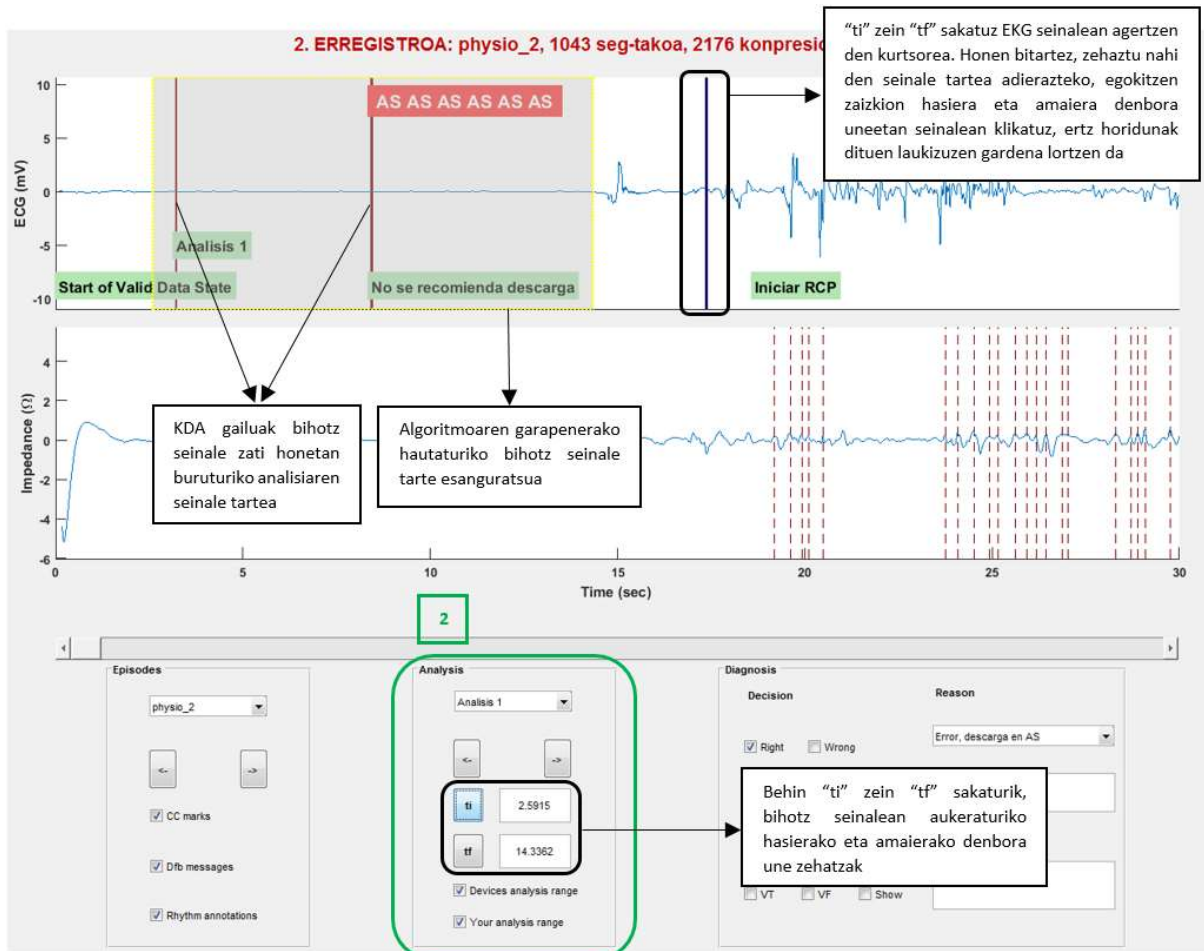
Irudia 11. Analisisetara bideraturiko bigarren azpi-panelak eskuragarri dituen funtzionalitate ezberdinen laburpena

Bigarren azpi-panel honek erakusten dituen funtzionalitate ezberdinen erabilpenaren adibide bat Irudia 12an azaltzen da. Bigarren erregistroari dagokion paziente berdinarekin jarraituz, alde batetik, lerro gorri bertikalekin KDA gailuak gauzaturiko bihotz seinalearen analisi tarte zehazten da. Bestetik, ertzak horiak dituen laukizuzen gardenak, aztertutako eta algoritmoaren garapenerako hautaturiko seinale tarte nabarmentzen da, hau 10 segundo ingurukoa gutxienez izatea gomendatzen delarik. Kasu honetan, KDA gailuak buruturiko seinalearen analisi tarte 3 eta 9 segundoen artean dagoela ikus daiteke; geuk zehaztutako analisisirako eta sailkatzaile automatikorako seinale tarte, kontrara, 2.5 eta 14 segundoen artean dago, horrela, gutxi gorabehera 12 segundoko seinale tarte adierazgarria lortzen delarik. Gainera, azkeneko balio hauek lehen aipaturiko *Edit Text* elementuetan ere adierazita ikus daitezke.

Algoritmoaren garapenerako bigarren fasean sortu behar den datu-basea osatzeko ezinbestekoak diren seinaleen tarte esanguratsuenak eta nabarmenenak biltzeko, tarte hauen denbora uneak biltzen dituen arraya bat gordetzea beharrezkoa izan da. Aipaturiko nahitaezko analisi tarteen uneak gordetzen dituen arraya, beraz, *physio_ann.mat* datu-basea osatzen duten egokitzen zaion erregistroetako bakoitzeko egituretan atzitu dira. Horrela, *t_analysis* deituriko eremu gehigarri honekin bat, erregistroen egiturek hartzen duten azkenengo itxura jarraian dagoen Taula 11an eskuragarri dago.

Taula 11. Pazienteen EKG seinaleen *physio_ann.mat* datu-baseko erregistroen egitura seinalearen azterketan hautaturiko seinale analisi tarteen denbora uneak biltzen dituen arraya gehiturik

Izena	Datu mota	Azalpena
reg_name	string	Paziente bakoitzari dagokion erregistroaren izena
ann	array<cell>	KDA gailuak pazientearen EKG seinalean adierazitako eta grabaturiko mezuak
t_ann	array<double>	KDA gailuak adierazitako mezuen t segundo aldiuneak
t_comp	array<double>	Bihotz-biriketako berpizte konpresio bakoitza burutu den t segundo aldiuneak
t_analysis	array<double>	Aurrerago algoritmoaren garapenerako sortzen den datu-basea osatzen duten azterturiko seinalearen analisi ezberdinetan aukeraturiko seinale denbora tarte uneak



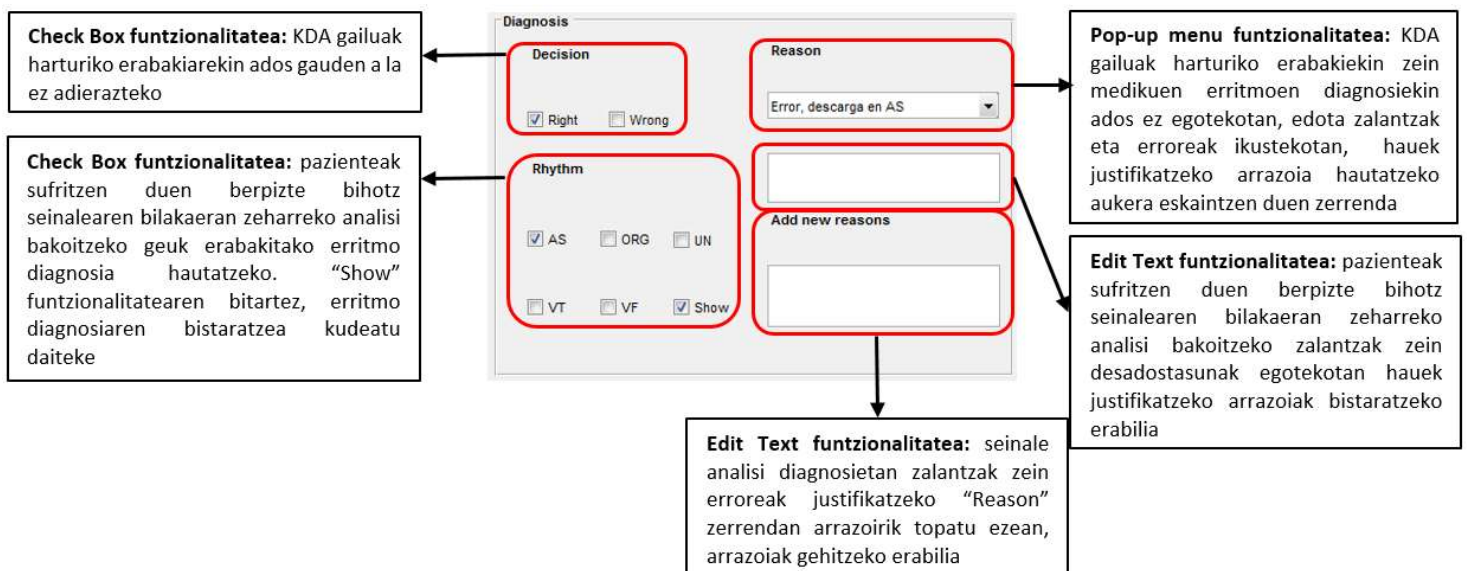
Irudia 12. Analisietara bideraturiko bigarren azpi-paneleko funtzionalitateak aktibo egonik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren azterlanaren emaitza gehigarriak

Plataformak eskaintzen dituen aipaturiko funtzionalitate eta elementu ezberdinek, seinaleen analisi sakon eta eraginkorrago bat egiteko gaitasuna ematen dute. Orain arte, KDAk bere aldetik buruturiko analisi tarte bakoitzeko harturiko erabakia ezagutzeko aukera ematen duela badakigu; hots, KDA gailuak pazientearen bihotz seinalearen analisiaren ondoren, honek momentuan pairatzen duen erritmo erabakiaren araberako terapia ezagutzea ahalbidetzen du, seinale tartea jasaten ari duen erritmoa hilgarria edo hilgarria ez denaren arabera gaixoarengan deskarga elektrikoa eraginez. Ez hori bakarrik, KDA gailuaren seinalearen bilakaeraren analisian zehar suertaturiko egoera ezberdinei buruzko mezuak ere ikusgai daude. Horrez gain, EKG seinalea, baita oso lagungarria izaten den TTI inpedantzia seinalea ere, eskuragarri daude, azken honetan, BBB maniobren ondoriozko interferentziak bistartzeko aukera izanik. Gainera, analisi tarte bakoitzeko amaierako erabakian asko laguntzen duten alorrean adituak diren sei medikuen erritmoen diagnosis ezagutzeko aukera ere badago.

Hirugarren eta azkenengo azpi-panela, diagnosira bideraturikoa hain zuzen ere, bihotz seinaleen analisian eta erritmoen erabakietan laguntzeko erabilia da. Hau da, funtzionalitate multzo honen

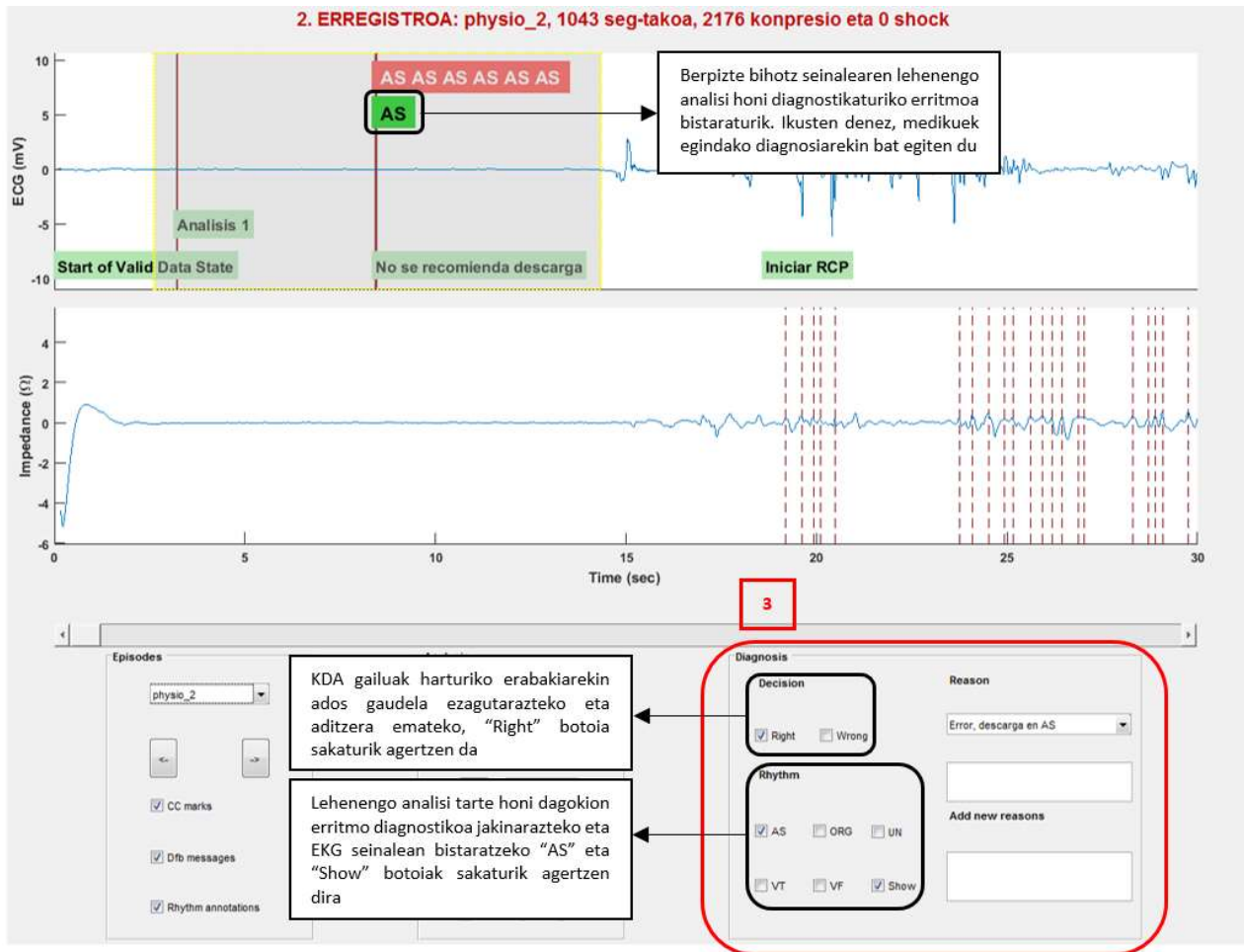
bitartez, algoritmo sailkatzaile automatiko baten garapenerako ezinbestekoa den datu-base baten osaketarako beharrezko seinale tarteen erritmo erabakiak kudeatzea ahalbidetzen da.

Deskribaturiko funtzionalitate hauetan guztietan oinarriturik, baita seinaleen analisisian funtsezkoak eta aurrerago sakonkiago azalduko diren oinarritzko irizpideetan oinarriturik ere, azkeneko panel honek hurrengo hiru funtzionalitateak eskaintzen ditu: alde batetik, KDAk buruturiko analisiaren ondoren, gailuak erdietsitako erabakiarekin bat gatozen ala ez jakinarazteko eta adierazteko, “Right” eta “Wrong” deituriko bi *Check Box* eskuragarri daude. Horrez gain, KDA gailuak gaixoaren EKG seinalearen bilakaeran zehar gauzaturiko analisi bakoitzeko jasaten duen geure erritmo erabakia adierazteko gaitasuna ematen duten bost *Check Box* eskuragarri daude; hots, pazienteak momentuan erritmo desfibrilagarriarik (FB, TB), erritmo ez desfibrilagarriarik (AS, ORG) zein zalantzazko erritmorik pairatzen duen adierazi daiteke. Gainera, geure erritmo diagnosis, adituak diren 6 medikuen erritmoen diagnosisekin bat, EKG seinalean bistaratzeko funtzionalitatea ere eskuragarri dago, hau ere “Show” deituriko *Check Box* baten bitartez kudeatua. Panelak eskaintzen duen hirugarren funtzionalitatea, KDAk harturiko erabakiarekin zein medikuntza esparruan adituak direnekin ados ez egotekotan, edota erritmoa erabakitzerakoan zalantzak izatekotan, hauek justifikatzeko arrazoi ezberdinak aurkezten duen zerrenda luzea eskuragarri duen *Pop-up Menu* funtzionalitatea da. Aurretiaz definituriko arrazoiaren artean, erabakitako tratamendu ekintzarekin huts egitea, shock egin duela esan baina KDA gailuaren karga baztertea eta txinpartik ez sortzea, erritmoak sailkatzeko konplexuak izatea, bitarteko erritmoak izatea eta analisisian konpresioen eragina izatea dira besteak beste. Aipaturiko zerrendan bilatzen dugun arrazoirik ez aurkitzekotan, “Add new reasons” deituriko *Edit Text* funtzionalitatearen bitartez guk geuk arrazoi bat gehitzeko gaitasuna ere eskaintzen du, gehituriko arrazoi hau, “Reason” zerrendan ere atzitzuz. Gainera, edozein analisi tarte batek harturiko erabakia justifikatzeko arrazoirik atxikiturik izatekotan, aipaturiko hau beste *Edit Text* funtzionalitate batean bistaratuko da. Beherago eskuragarri dagoen **Error! Reference source not found**.ak, aipaturiko hirugarren azpi-panelaren funtzionalitateak laburbiltzen ditu.



Irudia 13. Diagnosira eta analisisira bideraturiko hirugarren azpi-panelak eskuragarri dituen funtzionalitate ezberdinen laburpena

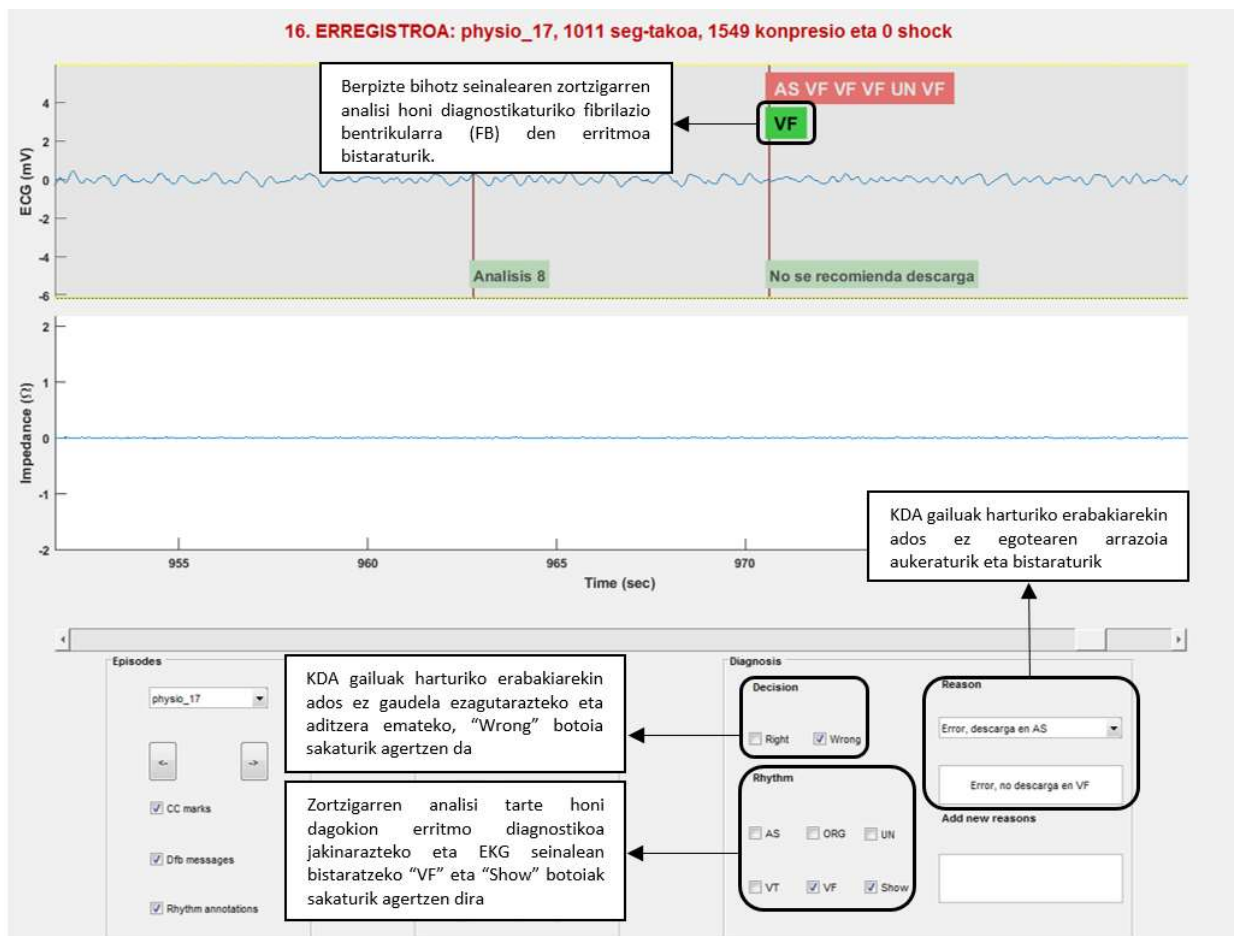
Aurretik deskribaturiko funtzionalitateen adibide berdinarekin jarraituz, Irudia 14ean 2.erregistro horri dagokion pazientearen berpizte bihotz seinalearen bilakaeraren lehenengo 30 segundoetako lehenengo analisi horretan, pazienteak deskarga elektrikorik behar ez duen KDA gailuak harturiko erabakiarekin ados gaudela adierazi da alde batetik; horregatik, "Decision" atazari dagokion "Right" Check Box-a sakaturik agertzen da. Halaber, seinale zati hori azterturik, mediku guztiekin bat, pazienteak asistolia (AS) jasaten duela erabaki da; izan ere, seinaleari so eginez, pazienteak miokardioan jarduera elektrikorik ez duela, hau da, erritmorik jasaten ez duela agerikoa da. Honekin loturik, beraz, "AS" Check Box-a ere sakaturik agertzen da. Ez hori bakarrik, "Rhythm" atalasearen barnean ere "Show" Check Box-a sakaturik agertzen da; honekin, EKG seinalean geuk erabakitako erritmo diagnosis gainerako mezuekin batera bistaratzea lortzen da. Bukatzeko, kasu konkretu honetan, KDA gailuaren erabakiarekin ados egoteagatik, eta seinaleari dagokion erritmo bilakaera bereziki argia izateagatik, "Reason" funtzionalitatea ez da ezertarako erabili, beharrezkoa izan ez baita.



Irudia 14. Diagnostikoetara bideraturiko hirugarren azpi-paneleko funtzionalitateak aktibo egonik, 2.erregistroari dagokion pazientearen lehenengo 30 segundoen bilakaeraren azterlanaren emaitza gehigarriak

Egoera ezberdinak aztertzearen, 16.erregistroari dagokion pazientearen berpizte bihotz seinalearen bilakaeran zeharreko zortzigarren analisia da Irudia 15ek aurkezten duen adibidea. Kasu honetan, erritmoa batere erregularra ez izateagatik FB diagnostikatzea ohikoena izaten da. Ikusten denez, sei medikuen diagnostikoetatik, lauk FB diagnostikatu dute, beste batek AS eta besteak zalantzazko erritmoa dela adieraziz. Fibrilazio bentrikularra askotan zalantzazkoa izatea zein asistolia erritmo batekin erratzea normala izaten da; hau, gailuak berak seinalean gehitu dezakeen zarata dela medio edota bular-sakaden ondorioz sortzen diren interferentziak direla eta gertatzen da. Garaturiko bihotz seinaleen kudeaketarako tresna honek, ordea, aurretik aipaturiko TTI inpedantzia seinalea ere eskuragarri du, aipaturiko interferentzia horien ispilu gisa jokatzeko duena. Aipaturiko abantaila horrekin jokatzuz, eta seinalearen azterketa gauzatuz, "Rhythm" atalasean adierazi bezalaxe, geure amaierako diagnostikoa gaixoak FB pairatzen duela da.

Kasu honetan, KDA gailuak buruturiko azterlanaren ondoren, honek plazaraturiko seinale egoeraren mezuen arabera, honen azkeneko erabakia pazienteak deskarga elektrikorik behar ez duela da. Geuk diagnostikaturikoa erritmo hilgarria denez gero, KDA gailuaren diagnostikoarekin ados ez gaudela "Wrong" adierazita jakinarazi, eta arrazoa ere bistaratu da: "Error, no descarga en VF"; hots, FB den erritmoan deskarga elektrikorik ez eragiteagatik, diagnostiko akastuna dela adierazi da.



Irudia 15. Diagnostikoetara bideraturiko hirugarren azpi-paneleko funtzionalitateak aktibo egonik, 16.erregistroari dagokion pazientearen zortzigarren analisi tartearen azterlanaren emaitza

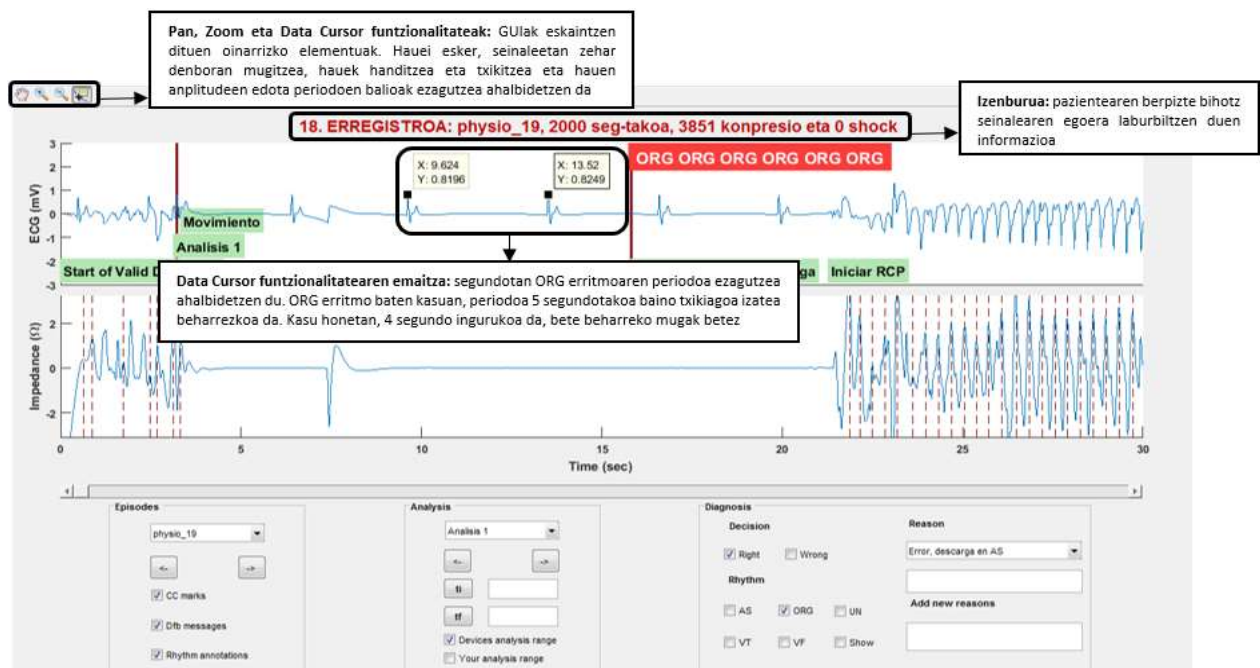
Aurrerago azalduko diren faseetan erabiliak izateko, diagnostikoetara bideraturiko hirugarren azpi-panel honek eskaintzen dituen funtzionalitate guztien bitartez seinalearen analisisian zehar lortzen diren eta algoritmoaren garapenerako informazio garrantzitsua ematen duten datu berriak biltegitratzea ezinbestekoa da guztiz. Kasu honetan, *resumenDatosExt.mat* taula egitura duen datu-basean atzitu dira, hurrengo itxura jaso duelarik:

Taula 12. Pazienteen EKG seinale bakoitzeko analisi bakoitzeko informazioa biltzen duen *resumenDatosExt.mat* datu-baseko taularen egitura seinalearen diagnostikoaren prozesuko datu berriak atxikiturik

Izena	Datu mota	Azalpena
Registro	string	Analisi tarte bakoitzari dagokion pazientearen erregistroaren izena
Fecha	double	Analisia egindako data
Analisis	double	Paziente bati egokitzen zaion seinaleak barnean hartzen dituen seinale bakoitzaren analisi zenbakia
DiagAED	double	KDA gailuak seinalearen azterketaren ondoren, pazienteari deskarga elektrikoa emateari buruz harturiko erabakia
RitmoCons	uint8	Medikuek adostutako (consenso) erritmo sailkapena
AnalisisOK	double	KDA gailuak harturiko erabakia zuzena den ala ez adierazten du
Comentarios	string	Medikuek anotazio prozesuan egindako iruzkinak
RitmoMedicos	array<uint8>	Alorrean adituak diren 6 mediku ezberdinen erritmo diagnosis
CommentsMedicos	array<cell>	Seinaleen analisisian medikuei suertaturiko egoerari buruzko informazioa ematen dituzten banakako iruzkinak
Decision	string	KDA gailuak harturiko erabakiarekin adostasuna zein desadostasuna adierazten du: "Right" edo "Wrong"
Rhythm	string	Seinaleak jasandako analisi bakoitzeko erritmo diagnostikoa adierazten du: AS, ORG, FB, TB, UN
Reason	string	KDA gailuaren erabakiarekin desadostasuna agertuz gero, zein errorerik edota zalantzarik izatekotan, hori justifikatzen duen arrazoia adierazten du

Garaturiko seinaleen kudeaketarako software plataforma honek eskaintzen dituen orain arte deskribaturiko oinarritzko funtzionalitate nagusiez gain, lagungarriak diren eta bihotz seinaleen bilakaeren analisisetan erraztasuna eskaintzen duten bigarren mailako funtzionalitateak ere eskuragarri ditu: *Zoom*, *Pan* eta *Data Cursor* funtzionalitateak dira hurrenez hurren. Hauek, GUIak berak zuzenean eskaintzen dituen funtzioak dira. *Zoom* funtzionalitateari esker, bihotz seinale tarteak handituz hauen azterketa xeheago, sakonago eta zehatzago bat lortzeko aukera ematen da. *Pan* funtzionalitatearen bitartez, paziente bakoitzaren bihotz eta inpedantzia seinaleetan denboran zehar mugitzeko gaitasuna ematen da, eta *Data Cursor* funtzionalitateaz baliatuz, seinalearen anplitude zehatzak edota periodoa ezagutzeko aukera ematen da. Hiru funtzionalitate hauetaz gain, *Pan* funtzionalitateak bezalaxe, EKG eta TTI seinaleetan zehar denboran aurrera eta atzera mugitzeko gaitasuna eskaintzen duen *Slider* funtzioa ere ahalbidetzen da. Bihotz-biriketakoa jasan duten paziente bakoitzeko plataformak uneoro bistaratzen duen izenburuak, gainera, berpizte bihotz seinalearen bilakaerak igarotako egoera ezberdinei buruzko laburpena eskaintzen du: pazienteari dagokion erregistro zenbakiaz eta

identifikatzaileaz informatzeaz gain, seinalearen guztizko analisiaren iraupena segundotan, pazienteak jasandako konpresio kopurua eta erritmo hilgarriak pairatzearen ondorioz jasandako shock kopurua ere aurkezten ditu.



Irudia 16. GUIak berez eskaintzen dituen Pan, Zoom eta Data Cursor funtzionalitateen erabilera aurkezten duen 18.erregistroari dagokion pazientearen adibidea

Irudia 16ari so eginez, 18.erregistroari dagokion pazientearen lehenengo analisiari dagokion seinalearen egoera ikusten da. Kasu honetan, begi bistaz, ORG erritmo antolatu normala jasaten duen gaixoa dela esan genezake; hala ere, ORG erritmo batek bete beharreko zenbait irizpide ere badaude. Hauen artean ezagunena, periodoa 5 segundotakoa baino handiagoa ezin daitekeela izan, hots, gutxienezko erritmoa 12 taupada minutukoa behar duela izan da. Aipaturiko “Data Cursor” elementuari esker, benetan aztertzen ari garen erritmo honen periodoa mugak betetzen dituen egiaztatzea ahalbidetzen du. Kasu honetan, periodoa 4 segundotakoa inguru dela baieztatu dezakegu, pazienteak benetan jasaten duen erritmoa ORG erritmo antolatua dela berretsiz. Era berean, seinalearen guztizko iraupena 2000 segundotakoa, jasotako konpresio kopurua 3851 eta deskarga elektronik pairatu ez duela ondorioztatu daiteke. Azkeneko baieztapen honekin batera, KDA gailuaren arabera, hasiera batean erritmo hilgarri jasan ez duela jakin daiteke.

Aurretik aipatu izan dugunez, Matlab-ek eskaintzen duen GUI interfazea izan da analisirako software kudeaketa plataformaren garapenerako erabili dena. Informazioa bistaratzea eta informazio hori kudeatzea ahalbidetzen duen interfaze bakarraren erabilera nahikoa izan da kasu honetan. Ildo honetatik, interfaze nagusian gehitutako elementu eta funtzionalitate guztiak .m script fitxategi nagusi batekin sinkronizaturik daudela, eta bertan, elementu bakoitzarekin zuzenki erlazioaturiko funtzioak definitzen direla esatea dago. Gainera, GUIak eskaintzen dituen elementu ezberdin bakoitzeko, guk helburu dugun funtzionalitatea burutzea lortzeko funtzioak moldatzea besterik ez dugu egin behar izan, eta beharrezkoa izatekotan, funtzio berriak sortu ere. Kudeaketa plataformak duen itxura ere, elementu bakoitza aurretik konfiguraturik lortu egin da.

Honekin, gauzatu beharreko hurrengo pausua den datu-baseak biltegiturik dituen 857 paziente ezberdinen seinaleen analisirako, anotaziorako eta segmentaziorako ezinbestekoa den aurretik eraiki eta garatu beharreko kudeaketa software plataformak eskaintzen dituen elementu eta funtzionalitate ezberdinen metodologia, deskribapena eta erabilera azaldu da. Eraikitako interfazea eskuragarri izanik, beraz, algoritmoaren garapenerako ezinbestekoak diren datuen prestaketarekin aurrera egin dezakegu.

8.1.2. Seinaleen analisisa, anotazioa eta segmentazioa

Machine Learning algoritmo baten garapenerako ebatzi nahi den arazoa ezaguturik, algoritmoa elikatzeko baliozkotasuna aurkezten duten datu multzoa ezinbestekoa da. Lortutako informazioaren kalitateak, bai eta kopuruak ere, sortuko dugun ereduaren amaierako funtzionamendu egokiarekin hertsiki loturik dago. Badakigunez, proiektu honetan burutuko den ML algoritmoa, gainbegiratutako algoritmo batean oinarritzen da; hau da, algoritmoa aurretik anotatutako datu multzo batez baliaturik ikasten du. Lehenago aipatu denez, kasu honetan, ospitaletik kanpoko 857 pazienteen erregistroetan oinarritutako datu-basea eskuragarri dugu. Datu-baseak eskaintzen duen informazioa, ordea, ikasketa automatikoa oinarri hartzen duen algoritmoaren garapenerako beharrezko anotatutako datuak ez ditu biltzen. Hori horrela, algoritmoa garatzeko erabiliko diren datuak prestatzeko, datu-baseak biltzen dituen 857 erregistroen analisisa egitea eta hauen anotazioa derrigorrezkoa bihurtzen da.

Aurretik deskribaturiko eta sorturiko seinaleen kudeaketarako eta analisirako tresna baliatuz, pazienteen berpizte bihotz seinaleen banakako anotazioa burutu da. Horretarako, gaixoen EKG eta TTI seinaleak aztertu, tresnak berak bistaratzen dituen sistema kardiobaskularraren eta arnas sistemaren medikuntza esparruan sei adituen erritmo anotazioetan oinarritu, eta orokorrean, medikuntza eremu honetan ezagutu eta jarraitu beharreko zenbait irizpide ere kontuan hartu dira.

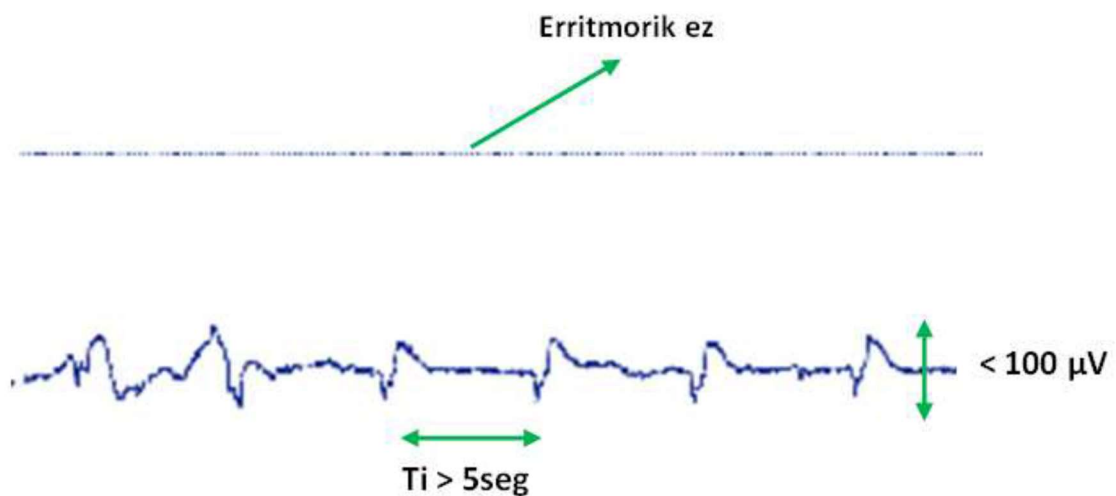
Azpi-atal honetan, beraz, pazienteen seinaleen banakako analisirako EKG eta TTI seinaleek ematen duten informazioan oinarritzeaz gain, jarraitutako irizpideak zeintzuk diren aipatuko dira. Azalduko diren irizpide hauek, datu-baseak erregistraturik dituen sei medikuek erritmo anotazioak egiteko jarraitzen dituztela ere aipatzekoa da.

Berpizte bihotz erritmoak, orokorrean, 5 kategoria ezberdinetan sailkatzen dira: asistolia (AS), pultsua aurkezten duen eta pultsurik gabeko antolatutako erritmoa, fibrilazio bentrikularra (FB) eta takikardia bentrikularra (TB) hurrenez hurren. Antolatutako erritmoekin loturiko pultsuaren detekzioak, ordea, zehaztasunean irabazteko, EKG seinaleaz gain, beste seinale batzuen beharra dauka ere; horretarako, pazienteak berezko pultsua duen ala ez datu-baseak jasotzea ezinbestekoa da, medikuek pazienteak tratatzen ari direnean anotazio horiek egitea beharrezkoa izanik. Gure kasuan, horrelakorik egitea ezinezkoa da, seinale nagusiak biltzen dituen KDA desfibrilagailu automatikoa medikuen menpe ez, eta independenteki jokatzen baitu. Aipatutako hau dela medio, sortuko den erritmoen sailkatzaile automatikoak, bost erritmo ez, 4 erritmo baino ez ditu bereiziko: asistolia (AS), antolatutako erritmo bakuna (ORG), fibrilazio bentrikularra (FB) eta takikardia bentrikularra (TB). Horrez gain, erritmo bat diagnostikatzerakoan zalantzak suertatzekotan, "UN" deituriko erritmo marka ere eskuragarri dago. Jarraian, erritmoen irizpide bereizgarrien azalpena banan-bana deskribatuko da:

Asistolia (AS):

AS diagnostikatuko da baldin eta bihotz seinaleak erritmorik ez badu edo erritmo antolatuaren itxura izatekotan, Irudia 17ko bigarren seinalea, periodoa (Ti) 5 segundo baino handiagoa bada, anplitudea $100\mu\text{V}$ baino txikiagoa bada eta bihotz erritmoa 12 taupada minutukoa baino txikiagoa bada.

Asistolia diagnostikatu aurretik, erritmoa ezkatututako beste erritmo bat ez den baieztatu behar da; kasu honetan, fibrilazio bentrikular leun batekin (FB leuna) es nahastea egiaztatu behar da. Gainera, pazienteak azkeneko erritmo hau jasanez gero, berpizte arrakastatsua lortzea konplexua da. Nahiz eta asistolia eta FB leun baten artean bereizitasunak ezberdintzeko zailtasunak egon, erritmoen arteko alde minimo hauek ere hautematea oso garrantzitsua da, asistolia egoera baten aurrean gaixoarengan deskarga elektrikoa eragitea arriskutsua oso gertatu baitaiteke. Irudia 18ak FB leun erritmo baten adibidea aurkezten du.



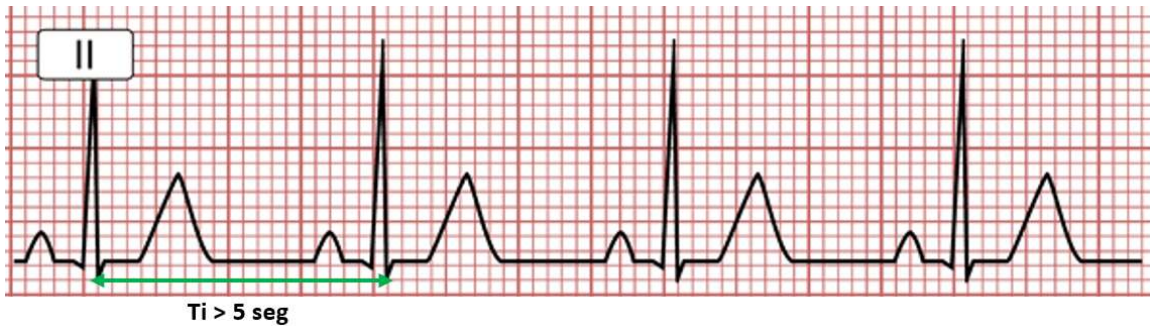
Irudia 17. Asistolia (AS) jasaten ari den pazientearen EKG bihotz seinalea



Irudia 18. Fibrilazio bentrikular leuna (FB leuna) jasaten ari den pazientearen EKG bihotz seinalea [65]

Antolatutako erritmoa (ORG):

ORG diagnostikatuko da baldin eta erritmoa erregularra eta periodikoa bada, periodoa (Ti) 5 segundo baino txikiagoa bada eta gutxienezko erritmoa 12 taupada minutukoa bada.



Irudia 19. Erritmo antolatua (ORG) jasaten ari den pazientearen EKG bihotz seinalea [66]

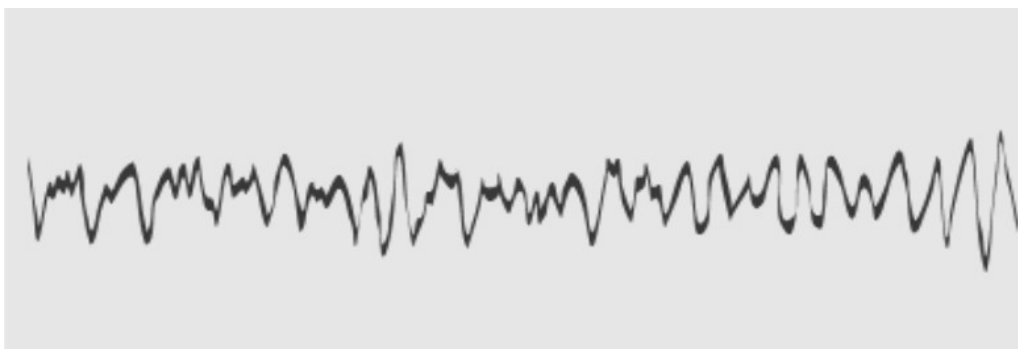
Fibrilazio bentrikularra (FB):

FB diagnostikatu da baldin eta, koordinaziorik gabeko jarduera elektrikoa jasanez, erritmoa erregularra ez bada eta anplitudea $200\mu\text{V}$ ingurukoa bada.



Irudia 20. Fibrilazio bentrikularra (FB) jasaten ari den pazientearen EKG bihotz seinalea [67]

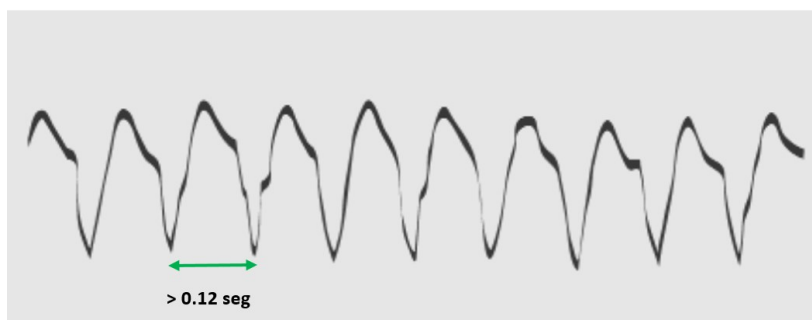
Seinale irregularrak anplitude handiagoa aurkeztea ere posiblea da, fibrilazio bentrikular honi sendoa deitzen zaiolarik. Honako arritmia, desfibrilazio goiztiarrarekin erraz zuzendu daitekeen erritmoa da. Honen adibidea Irudia 21ean ikus daiteke.



Irudia 21. Anplitude handiagoa aurkezten duen fibrilazio bentrikular mardula (FB mardula) jasaten ari den pazientearen EKG bihotz seinalea. Uhinen anplitude handiagoa, tamainaz, itxuraz aldakorrek, baita erritmo aldakorrek ere aurkeztuz, jarduera elektriko bentrikular kaotikoa ikusten da [68]

Takikardia bentrikularra (TB):

TB diagnostikatuko da baldin eta erregularra eta azkarra bada. Gutxienez, 150 taupada minutuko erritmoa jasan behar du. Horrez gain, QRS konplexuen zabalera, salbuespenak salbuespen, gutxienez 0.12 segundotakoak izan behar dira, Irudia 22an agertzen denez.



Irudia 22. Takikardia bentricularra (TB) jasaten ari den pazientearen EKG bihotz seinalea. Erritmoa erregularra, 158 taupada minutuko maiztasunarekin eta QRS konplexuekin aurkezten da [68]

Askotan TB geldo baten aurrean aurkitzea ere posiblea da. Kasu hauetan, erritmoa 60 eta 100 taupada minutuko maiztasunen artean ibiltzen da gutxi gorabehera. QRS konplexuen zabalera > 0.12 segundo ingurukoa izaten jarraitzen du.



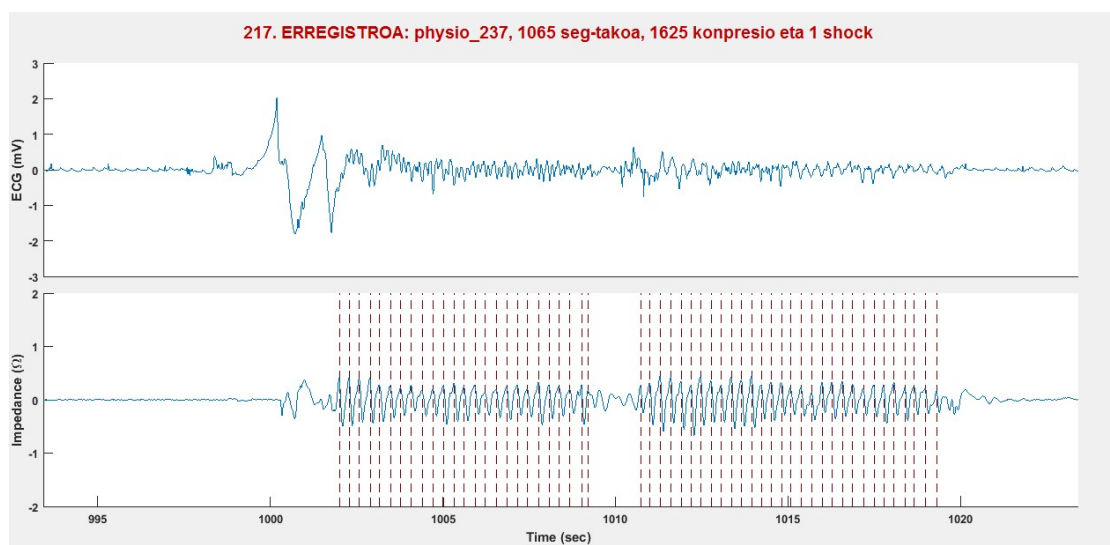
Irudia 23. Takikardia bentricular geldoa (TB geldoa) jasaten ari den pazientearen EKG bihotz seinalea

Zalantzako erritmoa (UN):

Zalantzako erritmo bat diagnostikatzearen oinarriko arrazoen artean Taula 13an adierazitakoak aurki daitezke. Horrez gain, erritmoa zuzenean argia ez bada ere zalantzako gisa anotatu daiteke ere.

Taula 13. Sailkatzeko konplexuak diren erritmoen oinarriko arrazoiak

AS/ORG anplitudeen arteko muga	FB/TB muga
AS/ORG maiztasunen arteko muga	FB/AS muga
AS/ORG anplitudeen/maiztasunen arteko muga	TB/ORG muga



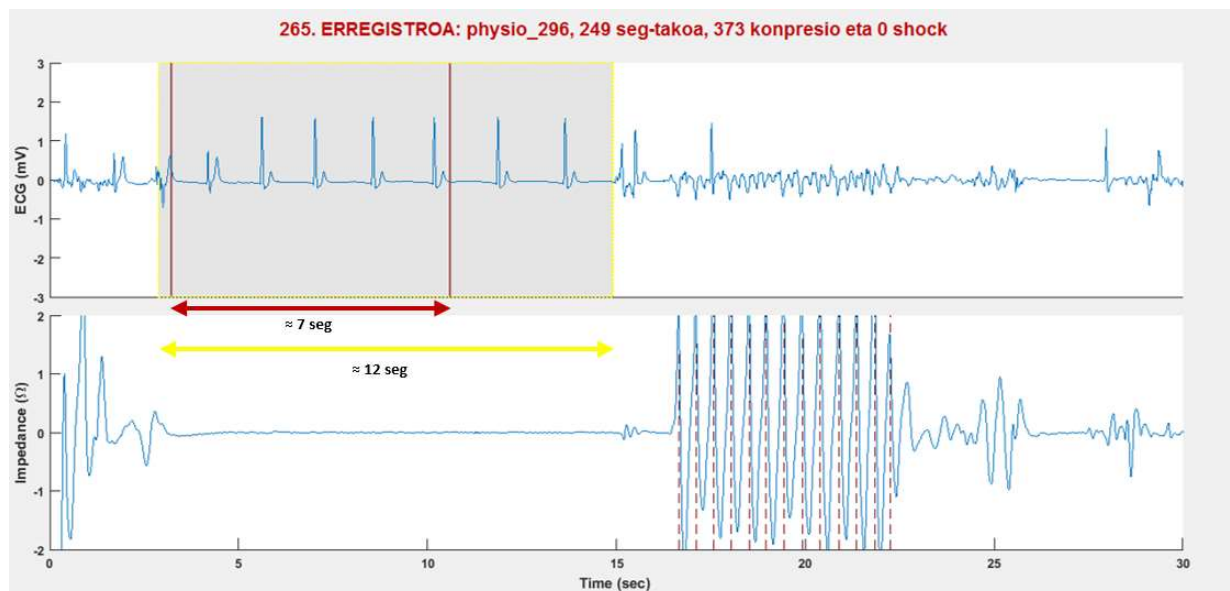
Irudia 24. Zalantzako erritmoa (UN) aurkezten duen 217.erregistroari dagokion pazientearen bihotz seinale tarte baten analisisa. Kasu honetan, begirada bakun azkar batekin, EKG seinaleak aurkezten duen erritmoa nahasia dela esan genezake. TTI inpedantziari so eginez, ordea, interferentziak eta konpresioak aurkezten dituela ikus genezake, honek, EKG seinalean bortizki eragiten duelarik. Honekin, hortaz, seinalea garbia eta argia ez dela, eta zalantzako erritmo gisa sailkatzea erabakitzen dela ondoriozta daiteke

Zalantzako erritmo baten adibide gisa, Irudia 24k 217.erregistroari dagokion pazientearen analisia aurkezten du.

Irizpide hauek guztiak kontuan harturik, baita EKG eta TTI seinaleen analisi sakona eginez ere, pazienteen banakako berpizte bihotz seinaleen azterketa aberatsa egiteko aukera ematen da, horrela, seinale bakoitzeko analisi guztien anotazio egoki bat eginez.

Erritmoen anotazio zuzena eginaz, analisi bakoitzeko seinaleen segmentazioak ere garrantzia du. Funtsean, berpizte bihotz seinaleen bilakaera osoan zeharreko analisi bakoitzeko, benetan adierazgarriak eta esanguratsuak diren seinale tarteak lortu nahi dira. Horrez gain, tarte hauen gutxienezko iraupena 10 segundotakoa izatea ere gomendatzen da. Era horretan, hurrengo fasean algoritmoaren garapenerako erabiliko diren laginen datu-basearen sorrerarako, informazio adierazgarria bakarrik aurkezten duten, eta ez zarata zein interferentziak direla eta kaltetuak izan diren, seinale-tarteak kontuan hartzea errazagoa izango da, osotasunean, datu multzo garbi bat erdietsiz.

Segmentaziora bideraturiko adibide gisa, Irudia 25ak 265.erregistroari dagokion pazientearen seinale tarte baten analisiaren adibidea aurkezten du. Irudiari so eginez, KDA gailuak bere aldetik independenteki buruturiko analisi tarteak, 7 segundo inguru iraun dituela ikus dezakegu. Aurretik esan bezala, analisirako seinale tarte gutxienez 10 segundotakoa izatea gomendatzen da, honek seinaleak lantzeko eta kudeatzeko erraztasuna eskaintzen duelako, bai eta informazio esanguratsu gehiago hornitzen duelako. Gainera, algoritmoaren garapenerako erabili diren seinale segmentuak 4 eta 8 segundotakoak izan dira. Aipatutako honengatik guztiarengatik, EKG seinalearen egoera azterturik, aukeratutako seinale segmentua 12 segundotakoa izatea aukeratu da.



Irudia 25. Seinaleen segmentaziora bideraturiko 265.erregistroari dagokion pazientearen seinale tarte baten analisia

Datu-basea osatzen duten 857 erregistroen seinaleen banakako analisen anotazioak eta segmentazioak buruturik, berpizte bihotz seinaleen sailkapen automatikorako algoritmoaren

garapenerako beharrezko den datu-basea sortzeko datu adierazgarri eta esanguratsuak prestatu eta lortu dira. Erdietsitako datu hauekin guztiekin, beraz, 2.fasearen helburu den algoritmoaren garapenean oinarrituko diren seinaleen datu-basea sortzea ahalbidetzen da.

8.2. 2-FASEA: Algoritmoaren garapenean oinarrituko datu-basearen sorrera

Proiektuaren bigarren fase honetan, berpizte bihotz erritmoen sailkatzaile automatikoaren garapenerako beharrezkoa den datu-basea sortu da. Aurreko atalean buruturiko seinaleen banakako analisi ezberdinen azterketan lortutako erritmoen anotazio eta segmentazioak baliaturik, algoritmorako nahitaezkoa, adierazgarria eta esanguratsua den informazioaz osatu da *db_analysis.mat* izendatutako datu-basea. Prozesu honetarako, Matlab-eko beste .m script fitxategi bat erabili da, kasu honetan *createDDBB.m* deitutakoa.

Datu-base honen sarrera bakoitza, erregistro bati dagokion paziente baten EKG seinalearen eboluzioan zehar buruturiko analisisetako bakoitzari egokitzen zaio, guztira 3803 sarrera lortuz. Aurreko atalean aipaturiko oinarritzko irizpideak betetzen ez dituzten EKG analisi segmentuak, zalantzarikoa seinale segmentuak, bai eta zaratatsuak eta kalitate baxuko EKG analisi segmentuak ere, atal honetan baztertuak direla ere azpimarratzekoa da. Honen xedea, erritmo ezberdinak benetan ezaugarritzen dituzten seinale segmentu garbiak baino ez izatea da, sortu nahi den algoritmoaren eraginkortasuna eta emaitza gaitasuna ahalik eta kalitate handienekoa izateko besteak beste. Datu-basea osatzen duten 3803 egituretako bakoitzak, beraz, hurrengo itxura dute:

Taula 14. Algoritmoaren garapenerako nahitaezkoa den db_analysis.mat datu-basearen egitura

Izena	Datu mota	Azalpena
reg_name	string	Analisi zehatz bati dagokion erregistro originalaren izena: "physio_1", "physio_2"...
c_analysis	double	KDA buruturiko analisi zenbakia adierazten du: 1, 2, 3 ...
int_aed	array<double>	KDAak buruturiko analisisirako erabilitako seinale tartea denbora unitatetan
int_cut	array<double>	Geuk GUI plataformaren bitartez markaturiko seinale analisi tarte esanguratsua denbora unitatetan
fs	double	Erabilitako laginketa maiztasuna
r_type	string	Analisi tartean erabakitako erritmo mota bateratua
s_ecg	array<double>	Geuk GUI plataformaren bitartez markaturiko seinale analisi tartearen (int_cut) EKG seinalearen laginen balioak

Lorturiko datu-baseak, esan bezala, guztira 4 berpizte erritmo (AS, ORG, TB, FB) bereizten dituen 3803 seinale segmentu biltzen ditu. Hautaturiko seinale segmentu bakoitzak 10-20 segundo inguru irauten ditu, aurrerago, algoritmoaren garapenerako 4 segundotako eta 8 segundotako segmentu egokiak aukeratzeko gaitasuna izateko. Horrez gain, datu-baseak aurkezten duen erritmoen proportzioak desoreka dezentea erakusten du: 3803 erritmoetatik, 2009 seinale segmentu AS erritmoari egokitzen zaizkio, datu-baseak barnean hartzen dituen laginen erdia baino zertxobait gehiago izanik. Hilgarriak ez diren erritmoekin jarraituz, erritmoen 1334ek ORG erritmoari egiten diote erreferentzia (datu-basearen gutxi gorabeherako %35). Gelditzen diren desfibrilagarriak diren erritmoei dagokienez, hau da, FB eta TB erritmoei dagokienez, hauei datu-baseak barnean hartzen dituen seinale segmentuetatik 441 eta 19 segmentu baino ez diote erreferentzia egiten, hurrenez hurren. Agerikoa denez, datu-baseak aurkezten duen erritmoen arteko desoreka argia da guztiz,

algoritmo errentagarrienaren aukeraketarako ebaluazioan kontuan hartzekoa dena. Horrez gain, datu-baseko seinale segmentu guztien laginketa maiztasuna 250 Hz-takoa dela aipatzekoa da. Jarraian, aipatutakoa laburbiltzen duen taula aurkezten da:

Taula 15. Datu-basearen berpizte bihotz erritmo klaseen proportzioa

Erritmo klase kopurua			
Erritmo ez desfibrilagarriak		Erritmo desfibrilagarriak	
AS	ORG	TB	FB
2009	1334	19	441

Datu-basea eta honen egoera ezaguturik, proiektuaren mamira egin dezakegu salto; hau da, helburu duen berpizte bihotz erritmoen sailkatzaile automatikoa sortzeko algoritmoaren garapenarekin has gaitzke, horrela, bihotz-biriketako berpizte egoeren aurrean aplikatu beharreko metodo eta teknika ezberdinen eraginkortasuna eta kalitatea, baita metodo eta teknika horiek hobetzera bideraturiko ikerketa oro sustatzea lortu nahi delarik.

8.3. 3-FASEA: Algoritmoaren garapena, ebaluaketa eta hautaketa

Hirugarren fase honetan, Machine Learning ikasketa automatikoa oinarritutako gainbegiratutako algoritmo baten garapenerako jarraitu beharreko prozesua da deskribatuko dena. Kasu honetan, berpizte bihotz erritmoen sailkatzaile automatikoa eraiki nahi da, bihotz-biriketako geldialdiekin loturiko medikuntza esparruko etorkizuneko lanetan, terapietan, tratamenduetan, diagnostikoetan zein ikerketetan erraztasunak, hobekuntzak eta laguntza eskaintzeko.

Prozesua, bereizitako azpi-atal ezberdinetan banatuta dago. Lehendabizi, algoritmoaren entrenamendu eta ebaluaziorako erabiliko diren datuen aurretiazko prozesaketa gauzatu behar da, jarraian, datu-baseko seinaleak ezaugarritzen dituzten ezaugarrien erauzketa burutu behar delarik. Ondoren, askotan aipatu den *overfitting* arazoa ekiditeko, balioztatze gurutzatua aplikatu da, amaierako eredu ahalik eta eraginkorra izatea lortu nahi delarik. Seinaleen ezaugarri bereizgarriak lorturik, baita balioztatze gurutzatua aplikaturik ere, ikasketa automatikoa oinarritutako gainbegiratze ikasketan funtsatutako RF teknika aplikatzea da hurrengo pausua, algoritmoa entrenatuz alde batetik, eta balioztatuz bestetik. Eredua aplikaturik, lorturiko algoritmo ezberdinen errendimenduaren ebaluazioa da burutu beharreko azkeneko pausua, horrela, eraginkortasun eta fidagarritasun maila handiena eta onena aurkezten duen sailkatzailea erdietsiz.

8.3.1. Seinale segmentuen prozesaketa

Atal honetan, algoritmoaren garapen eraginkor baterako eta hurrengo faseetara egokitzeko azterturiko eta analizaturiko seinaleen bateratzea bilatzen da. Seinale elektrokardiografikoa jatorri ezberdinetatik eratorritako zaratagatik kaltetua dela jakina da; horrenbestez, sortzen duten interferentzia murrizteko seinalea beraren iragazketa burutzea ezinbestekoa da.

Kasu honetan, maila lau duen 0.5 eta 30 Hz arteko banda-paseko Butterworth iragazkia erabili da; izan ere, kanpoko desfibrilagailu automatikoez darabilten banda zabalera 0.5 eta 30 Hz bitartekoa izatea da ohikoena.

8.3.2. Ezaugarrien erauzketa

Ezaugarrien erauzketa prozesua, berpizte bihotz erritmoak elkarren artean bereiztea posiblea egiteko seinale ezberdinak ezaugarritzen dituzten ezaugarrien multzo bat lortzean datza. Honako zeregina, erritmoen sailkapenerako garrantzi handia hartzen duen prozesua da, aukeratutako ezaugarriak, aurrerago sailkatzailea entrenatzeko, baita klase mota ezagutzen ez duten seinale segmentu bakoitzeko aintzat hartu behar dituen atributuak zeintzuk diren jakinarazteko ere erabiliak baitira.

Ezaugarrien hautaketa, ikertzen ari garen esparruaren edota gaiaren araberakoa izaten da; era berean, ezaugarri batzuek sailkatzaileari ekarpen baliotsurik ez egitea, edota sailkatzailea nahastea posiblea da, beste egoera batzuetan, kontrara, erabilgarritasun eta balio handiak eskaintzen dutelarik.

Geure helburu nagusia, ahalik eta informaziorik esanguratsuen eta nabarmenena eskaintzen duten ezaugarri kopuru handiena lortzea da. Atributu hauen kopurua zenbat eta handiagoa izan, klasifikatzaileak orduan eta zehaztasun handiagoa erdietsi eta emaitzak errealitatera orduan eta gehiago hurbilduko dira. Seinaleak deskribatzen dituzten ezaugarri bereizgarriek, beraz, bereziki sailkatzaileen jardura asko hobetu, eta erreduantzia ere murrizten dute. Hala ere, aldagaiak gehitzen doazela, sailkatzailearen entrenamendu kostea handituz doaz, eta sailkatzaileak orokortze erroreak gauzatzeko arriskua asko handitu daiteke ere [69].

Kasu honetan, iragazitako 4 eta 8 segundotako 3803 EKG segmentuetan oinarriturik, 33 ezaugarri bereizgarri kalkulatu dira, emaitza gisa, banakako iraupen bakoitzerako 33x3803 dimentsioko matrizea lortuz. Kasu honetan, eta aurretik aipatu denez, EKG berpizte bihotz seinaleak bereizteko 4 talde ezberdinetan banatutako ezaugarriak hautatu dira: denbora domeinuan oinarritutako ezaugarriak, ezaugarri espektralak, denbora-maiztasun ezaugarriak eta konplexutasun ezaugarriak hain zuzen ere.

Seinale segmentu bakoitzaren prozesaketarako, baita 33 ezaugarri bereizgarri hauen erauzketarako, Matlab-eko *computeFeatures.m* izeneko .m script fitxategi bat erabili da. Fitxategian, alde batetik, paziente bakoitzaren identifikatzailea duen "patID" deituriko array bat definitu da. Horrez gain, aztertutako eta analizaturiko EKG seinale segmentu bakoitzeko erritmo erabakia barnean hartzen duen "y" izendaturiko beste array bat ere definitu da. Aipaturiko azkeneko array honek, seinale segmentu bakoitzeko erabakitako erritmo bateratua balio ezberdinen bitartez adierazita ditu: AS (0), ORG (1), TB (2) eta FB (3). Azkenik, lortutako seinale segmentu bakoitzeko 33 ezaugarriak 33x3803 dimentsiodun matrize batean biltegitatu dira, azkeneko hau "X" deituz. Ezaugarri hauen guztien kalkulua, banaka definituriko .m script fitxategi ezberdinetan definituriko funtzioen bitartez burutu direla aipatzekoa da. Hori horrela, "patID", "y" eta "X" elementuak gorde eta biltzen dituen bi .mat fitxategi sortu dira; bata, 4 segundotako iraupena duten EKG segmentuetan oinarritutakoa ("*featureDDBB_4.mat*"), eta bestea, 8 segundotako iraupena duten EKG segmentuetan oinarritutakoa ("*featureDDBB_8.mat*").

Algoritmoaren eraginkortasuna, efizientzia eta kalitatea ahalik eta onena izatea lortzeko helburuarekin, 4 eta 8 segundotako EKG segmentuekin lan egitea erabaki da; horrela, emaitzen arteko

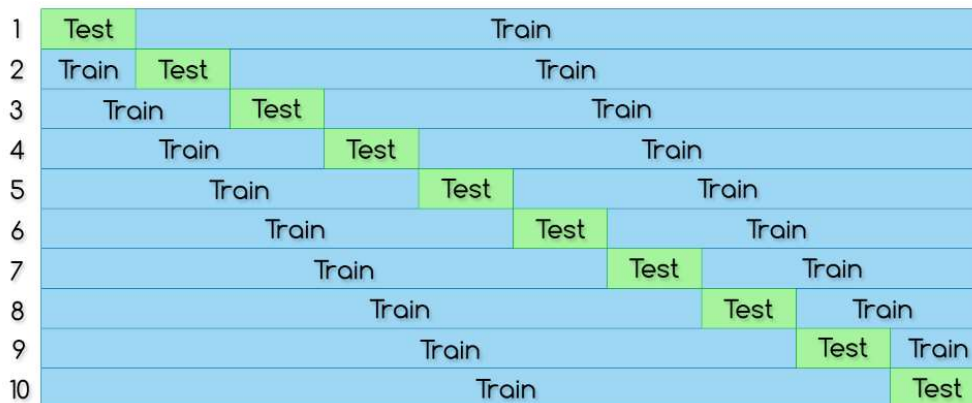
aldeak aztertuz, analizatuz eta ebaluatuz, bi seinale tarteen artean emaitzarik esanguratsuenak eta errentagarrienak zeinek ematen dituen ondorioztatu baitaiteke, betiere, berpizte bihotz erritmoen sailkatzaile automatiko onena erdiesteko xedearekin.

8.3.3. Balioztatze gurutzatua

Badakigunez, Machine Learning ikasketa automatikoak eskaintzen dituen algoritmoen gehiengoak *overfitting-a* jasateko joera izaten dute orokorrean, orokortu nahi den *helburu funtzioarekin* loturarik ez duten algoritmoaren entrenamenduan erabilitako laginen ezaugarri bereziki xeheetara doitzen baitira. Logikoaenez, ikasketa automatikoan funtsaturiko algoritmoek entrenamendu datuen baldintzak orokortzeko, baita ikasitakoa kasu berrietan iragartzeko eta aurrerateko ere, gaitasuna aurkeztu behar dute.

Kasu honetan, garatu den berpizte bihotz erritmoen sailkatzaile automatikoaren emaitzak aztertzerakoan faktore erabakigarri bat agertzen da: datuen banaketa, alde batetik, entrenamendu datuetan (*training set*) eta testeatze datuetan (*test set*), bestetik. Banaketa hau burutzen den eraren arabera, baita banaketa zenbat aldiz gauzatzen denaren arabera ere, garaturiko sailkatzailearen benetako zehaztasuna eta doitasuna definitzen da. Aipaturiko *overfitting* arazoari aurre egiteko, beraz, baita geure sailkatzailea ahalik eta zehatzena eta eraginkorra izateko ere, aipaturiko balioztatze gurutzatua da aplikatu dena.

Aurretik azalduenez, balioztatze gurutzatua deritzon metodoa entrenatzeko eta balioztatzeko daukagun datu multzoa k zatitan banatzean (k zati hauetako bakoitza *fold* gisa ezaguna) oinarritzen da, jarraian, prozesua k aldiz errepikatzen delarik. Prozesua errepikatzen den bakoitzean, balioztatze datu multzo gisa *fold* ezberdin bat erabiltzen da, gainerako datu multzoa entrenatzeko erabiltzen delarik. Modu horretan, gure sailkatzailearekiko buruturiko neurri guztiak, gauzaturiko exekuzio ororen batez bestekoa jorratuz kalkulatu dira. Honekin, sailkatzailearen efizientziarekiko balio are fidagarriagoak erdiestea lortzen da. Gure kasuan erabilitako k faktorearen balioa 10 izan da, grafikoki hurrengo irudiak aurkezten duelarik:



Irudia 26. Corpus-aren banaketa balioztatze gurutzatuko teknika erabilia $k=10$ balioarekin [69]

Aipaturiko $k=10$ baliodun balioztatze gurutzatuarekin, biltegituriko datu-baseko pazienteen %10a balioztatze datuetara eta %90a entrenamendu datuetara bideraturik doaz. Aipaturiko zatiketa hau pazienteko egiten da. Guztira, entrenamendu eta testeatze faseetan erabiliak izan diren seinale garbi eta errorerik zein zalantzarik gabeko erritmoak aurkezten dituzten 834 pazienteen EKG erregistroak eskuragarri izan ditugu. Balioztatze gurutzatua aplikaturik, pazienteen %10a, hots, 83 zein 84 paziente, balioztatze prozesurako erabili dira, gainerako 751 zein 750 pazienteak, hurrenez hurren, entrenamendu faserako erabili direlarik. Gainera, are eraginkortasun eta errentagarritasun altuago bat lortzeko helburuarekin, deskribaturiko $k=10$ balioztatze gurutzatu prozesu osoa nahi beste aldiz errepikatzea ere posiblea da; zenbat eta errepikapen kopurua altuagoa izan, makinaren prozesatze denbora ere altuagoa izango dela agerikoa da. Horregatik, geure kasuan, prozesua 100 aldiz errepikatzea aukeratu da.

Entrenamendu eta testeatze partizio orekatu bat lortzea ere garrantzitsua da. Hori dela bide, datu-base osoak barnean hartzen dituen erritmo bateratuen proportzioak kontuan harturik, entrenamendu prozesurako datuetan antzeko proportzio bat eskuratzea izan dugu helburu. Entrenamendurako erabiliko diren erritmo bakoitzaren proportzioa jatorrizko datu-baseko erritmoen proportzioetara zenbat eta hurbilago egon, algoritmo-sailkatzaileen emaitzen balioa aberatsagoa, ziurragoa, fidagarriagoa zein zehatzagoa izatea lortuko da. Horretarako, erritmo-proportzio ezberdinetako hainbat saiakera burutuz, lortutako azkeneko emaitzarik altuena, jatorrizko datu-baseko proportzioen %85a betetzea izan da; hots, datu-base originaleko asistolia (AS) erritmoen guztizko kopurua ezagututa, kopuru horren gutxienezko %85a entrenamendu faserako erabili da, gainerako erritmoekin proportzio berdina ere mantenduz. Proportzio altuago bat lortzea ezinezkoa suertatu da, programaturiko fitxategia begizta infinitu batean sartu, eta bertatik irtetea ezinezkoa bilakatzen baitzen.

Funtsezko balioztatze gurutzatuaren zatiketa egoera hau burutzeko, Matlab-ek eskaintzen duen beste .m script fitxategi bat erabili da, kasu honetan, *cv_partition.m* izendatua. Fitxategian bertan, aurretik sorturiko *db_analysis.mat* datu-baseak eskaintzen duen informazioaz baliaturik, baita seinaleak deskribatzen dituzten ezaugarri bereizgarrien erauzketan "*featureDDBB_4.mat*" eta "*featureDDBB_8.mat*" fitxategietan biltegituriko analizaturiko seinale segmentu bakoitzeko erritmo bateratua gordeta duen "y" arraya-ko informazioa erabiliz ere, helburu dugun algoritmo eraginkorrago bat lortzeko eta deskribaturiko balioztatze gurutzatua aplikatzeko, agindu multzo batez osaturiko kodea idatzi da. Horrela, definituriko proportzioan $k=10$ balioztatze gurutzatua 100 aldiz errepikaturiko *fold* bakoitzeko entrenamendu eta balioztatze bakoitzak, bakoitzari egokituriko pazienteei dagozkien analizaturiko EKG erritmo bateratu segmentuak zeintzuk diren ezagutu daitezke.

cv_partition.m fitxategia exekutaturik, emaitza gisa lorturiko *partition.mat* fitxategian bilduriko 100 aldiz errepikaturiko partizioak, hurrengo taulak aurkezten duen itxurako egitura batean biltegituratu dira:

Taula 16. Balioztatze gurutzatua aplikaturik lorturiko emaitzak biltzen dituen "*partition.mat*" fitxategiko egituraren itxura

cv_part		Egiturak hartzen duen izen nagusia, guztizko kopurua 100 izanik
folds		Sorturiko 100 balioztatze gurutzatuko partizio bakoitzeko 10 <i>fold</i> -ak biltzen dituen elementua
train	test	Sorturiko 100 balioztatze gurutzatuko partizio bakoitzeko 10 <i>fold</i> -etako bakoitzetako entrenamendu eta balioztatze datuak
array<logical>		Datu mota

8.3.4. Algoritmoaren garapenerako entrenamendu eta balioztatze faseak

Fase honetan da benetan Machine Learning teknikak aplikatzen diren unea. Algoritmoaren garapenerako datuen prozesaketa eta prestaketa guztiz osaturik dagoela, lehendabizi, algoritmoaren entrenamendua aplikaturik, ikasketa automatikoan oinarrituriko algoritmoa bera elikatu, eta balioztatze prozesua aplikaturik, algoritmoak entrenamenduan ikasitakoa frogatu eta ebaluatu egiten da. Entrenamendu fasearen ideia nagusia, helarazten zaizkion datuetatik informazio baliagarria atera, eta ondoren iragarpenak egitea posible izatea da. Testeatze prozesuan zehar, eraikitako algoritmoa iragarpenak eta aurrean egiten zein zehatza eta eraginkorra den ebaluatzen da, baita honek aurkezten duen errendimendua ere.

Proiektu honek helburu duen berpizte bihotz erritmoen sailkatzaile automatikoaren garapenerako algoritmo eraginkor baten sorketarako Machine Learning ikasketa automatikoak eskaintzen duen Random Forest teknika erabili da; izan ere, metodo honek malgutasuna, erabiltzeko erraztasuna, abiadura eta doitasuna aurkezten baititu. Ikerketa proiektu hau, identifikazioan eta klase edota kategoria ezberdinen arteko banaketan erakusten duen zehaztasunean erabakigarria izateko, RF teknikan funtsaturiko sailkatzaile ezberdinen azterketan oinarritzen da.

Entrenamendu eta balioztatze faseen garapenerako, *computeClassifier.m* izeneko fitxategia erabili da. Bertan, Matlab-ek eskaintzen duen RF sailkatzailea sortzeko `TreeBagger` funtzioa definitu da; funtzioak berak sailkatzailea eraikitzeo eskaintzen dituen parametroen artean, bi dira geuk erabakitzeo gailendutakoak. Funtzioaren oinarritzko definizioa hurrengoa da:

```
Mdl=TreeBagger(NumTrees,X_train,Y_train,'Prior','Uniform','SplitCriterion',  
'deviance','PredictorSelection','curvature','MinLeafSize',nMinLeafSize);
```

ML teknika berdin batean funtsaturiko sailkatzaile ezberdinen errentagarritasuna eta errendimendua aztertzeo helburuarekin, bi parametroen araberrako sailkatzaileak eraikitzeo erabaki dela aipatu dugu: kasu honetan, funtzioan definiturik ikusten diren `NumTrees` eta `nMinLeafSize` elementuak dira definitu direnak hain zuzen ere. Zuhaitz kopurua zehazten duen parametroari (`NumTrees`) esker, algoritmoak basoan zenbat zuhaitz erabiliko dituen zehaztea ahalbidetzen da. Horrez gain, zuhaitz bakoitzak izan ditzakeen nodo kopurua adierazten duen `nMinLeafSize` elementua zehazteo gaitasunak, zuhaitza zeinen sakona izan daitekeen definitzeo ahalbidetzen du, hau da, zuhaitz nodo bakoitzeko behaketa kopuru minimoa adierazten du. Horrela, aipaturiko bi parametro hauen araberrako sailkatzaile ezberdinen emaitzen kalitatea konparaturik, algoritmorik eraginkorrena hautatzeo posiblea da. Proiektuan geuk probatu ditugun bi parametro hauen balio ezberdinak hurrengoak izan dira:

`NumTrees`: 25, 50, 75, 100, 125, 150, ..., 350

`nMinLeafSize`: 1, 3, 5, 10, 15, 20

Bi sarrera huez gain, `X_train` eta `Y_train` sarrerak ere definitzeo nahitaezkoak dira. `X_train` parametroak, behaketa bakoitzeko algoritmoaren garapenerako ezinbestekoak diren ezaugarrien

multzoa biltzen du, kasu honetan, entrenamendurako erabiltzen dugun seinale segmentu bakoitzeko ezaugarri bereizgarriak biltzen dituelarik. `Y_train` parametroak, behaketa bakoitzeko erritmo erabakiak biltzen ditu, hau da, seinale segmentu bakoitzari dagokion erritmo erreala dagokion zenbakizko balioa biltzen du; horrela, seinale segmentuek aurkezten dituzten ezaugarri bereizgarriak eta hauen banakako zenbakizko erritmo anotazio erreala ezaguturik, algoritmoak ikasteko gaitasuna hartzen du.

Random Forest algoritmoa burutzeko erabiltzen den `TreeBagger` funtzioa exekutaturik, honek RF algoritmoaren datu guztiak barnean hartzen dituen `Mdl` deituriko sailkatzaile eredu bat emaitza gisa bueltatzen du. Sailkatzaile ereduak eskaintzen duen informazioaz, baita balioztatze gurutzatua aplikaturik balioztatze faserako hautaturiko seinale segmentuei dagokien ezaugarri bereizgarriez baliaturik ere, iragarpenak egitea hurrengo pausua da; Matlab-ek, horretarako, `predict` funtzioa erabilgarri dauka, oinarritzko definizioa hurrengoa izanik:

```
[y_fit, scores] = predict (Mdl, X_test);
```

Ikusten denez, sarrera gisa, entrenatutako analisi diskriminatzailean oinarritutako sailkatzailearen ereduaren `Mdl`, eta balioztatze prozesurako egokituriko seinale segmentuen ezaugarri bereizgarriak gordetzen dituen `X_test` matrizea definitzen dira. Era berean, irteerako lehen argumentuak, `y_fit`, ebaluatzen duen behaketa bakoitzeko iragarpena, hots, aurreikuspena gordetzen du.

Aipatu denez, sailkatzailearen eraginkortasuna ahalik eta altuena izatea erdiesteko, algoritmoaren garapenerako paziente multzoari *k-fold* deituriko $k=10$ baliodun balioztatze gurutzatua aplikatu zaio; horrela, entrenatze eta balioztatze faseetarako paziente konbinazio ezberdinak aurkezten dituzten 10 partizio (*fold*) lortzea posiblea izan da. Orain arte deskribaturiko prozesu hau, hots, sailkatzaile ereduaren sortzearen eta iragarpenak egitearen prozesua, *fold* bakoitzeko burutu beharreko prozesua izan da. `TreeBagger` exekutatzen den bakoitzean, ondoren, balioztatze prozesua gauzatzeko `predict` funtzioan erabiltzen den *test fold* bakoitza, 10 aldiz doitzea beharrezkoa dela, eta horrela, ditugun behaketa guztiei dagokien iragarritako emaitzak biltzen dituen `y_fit` arraya guztiz osatzen dela nabarmentzekoa da.

Entrenamendu eta balioztatze faseak buruturik, algoritmoen ebaluazioa eta hautaketa da geratzen den azkeneko pausua.

8.3.5. Algoritmoen ebaluazioa eta hautaketa

Eraikitako algoritmo bakoitzeko ebaluazioa burutzeko, hainbat ikuspuntuetatik eraikitako erritmoen sailkatzailearen zehaztasuna eta eraginkortasuna erraz ebaluatzeko ahalmena eskaintzen duen nahasmendu matrizea da (ingelesez *Confusion Matrix*, *CM*) kalkulatu den hurrengo elementua. Horretarako, `confusionmat` izeneko funtzioa erabili da, bere definizioa jarraian aurkezten dena izanik:

```
C = confusionmat (y, yfit);
```

Ikusten denez, sarrera datu gisa `y` eta `yfit` parametroak hartzen ditu; lehenengo parametroak sailkatzaileak ebaluatutako datuen anotazio balio erreala biltzen ditu, bigarren parametroak sailkatzaileak iragarritakoak gordetzen dituelarik. Bi datu hauen konbinazio eta laguntzaz, beraz, aipaturiko CM matrizea sortzea posiblea da, kasu honetan, C aldagaian gordez. Matrize honek hartzen

duen itxura, aurreko 5.atalean algoritmoaren ebaluazio irizpideak deskribatzean, proiektu honen kasua den berpizte bihotz erritmoen sailkapenerako bistaraturiko Taula 1 kontingentzia taulak aurkezten du. CM matrizea taula eran bistaratzeak, beraz, garaturiko sistemak sailkatzen dituen klaseak modu simple batean gaizki sailkatzen dituen ala ez ikusteko laguntzen du bereziki.

Askotan errepikatu denez, eredu sailkatzailearen errendimendua ahalik eta eraginkorra izatea lortzeko, $k=10$ partiziodun balioztatze gurutzatua 100 aldiz errepikatu da; hori horrela, CM matrizearen kalkulua, aztertutako algoritmo kasu bakoitzerako, 100 aldiz errepikatu behar izan da.

Partizio bakoitzeko CM nahasmendu matrizea kalkulaturik, ebaluazio metriken kalkulua da sorturiko sailkatzaile bakoitzeko errendimendua eta doitasuna neurtzeko falta dena. Horretarako, 5.ataleko ebaluazio irizpideen deskribapenean azaldu zenez, CM matrizeak berak eskaintzen duen informazioan oinarrituta, klase zehatz bakoitzeko sentikortasuna eta iragartze balio positiboa neurtu dira alde batetik, eta bestetik, CM matrizea parametro bakar batera murrizteko *MulAcc* eta *UMS* balioak dira kalkulatu direnak ere. Dagoeneko azpimarratu denez, $k=10$ partiziodun balioztatze gurutzatua 100 aldiz errepikatzeagatik, CM matrizea ere 100 aldiz kalkulatu behar izan da, eta ondorioz, ebaluazio metrika bakoitza ere 100 aldiz kalkulatu izan da. Metrika hauen guztien kalkulua, *classifier.m* deituriko fitxategi batean burutu dira.

Sailkatzaile automatiko eraginkorra, errentagarriena eta zehatzena hautatzen dugula ziurtatzeko, zuhaitz kopuruaren (*NumTrees*) eta zuhaitzak duen nodo kopuruaren (*nMinLeafSize*) araberako konbinazio guztiak aztertu direla era jakina da. Hori dela eta, *NumTrees* eta *nMinLeafSize* parametroen araberako konbinazio guztietan oinarrituriko emaitza guztiak, *Results.mat* izendaturiko fitxategian gordetako egitura bakarrean bildu dira (Taula 17an eskuragarri). Gogoratu, kasu honetan, 4 eta 8 segundotako seinale segmentuekin lan egin dela; hori dela eta, emaitzen bi fitxategi lortu dira: bata 4 segundotako seinale segmentuei bideraturikoa, eta bestea, 8 segundotako seinale segmentuei bideraturikoa.

Emaitzak datu-base batean biltegiturik, hauen azterketa erraza eta, aldi berean, aberatsa eta eraginkorra izateko, grafikoki adierazi dira, horrela, algoritmoen emaitzen konparaketa ere erosoagoa eta arinagoa izateko. Lorturiko emaitzak grafikoki bistaratzeko, parametro bakoitzeko lorturiko 100 emaitzen 3 pertzentil kalkulatu dira: P_{25} , P_{50} eta P_{75} . Horrez gain, bakoitzak lorturiko maximoak eta minimoak ere kalkulatu dira. Pertzentilak, lorturiko emaitzak txikienetik handienara ordenatu direla, portzentaje batean oinarritutako aldagai zehatz baten balioa itzultzen du; hots, behaketa lagin talde baten ehuneko horren laginen balioek, pertzentilak itzultitako balio zehatz horrek hartzen duen balioa baino baxuagoak direla aditzera ematen du. Pertzentil hauen kalkulurako *computePercentile.m* fitxategia erabili da, kasu honetan, Matlab-ek eskaintzen duen *prctile* funtzioa aplikaturik:

```
Y = prctile (X, p);
```

Funtzioaren definizioari so eginez, X elementuak, pertzentilak kalkulatu nahi diren datuen multzoa adierazten du, eta p elementuak, neurtu nahi diren pertzentilen ehunekoa adierazten du. Gainera, emaitzen bistaratzeko Matlab-ek eskuragarri duen *Box Plot* funtzioa hartu da oinarri. Aipaturiko pertzentilen emaitzak, *Results.mat* fitxategian atzitu egin dira (Taula 18an eskuragarri).

Taula 17. “Results.mat” fitxategian biltegituriko emaitzen egitura

Izena	Datu mota	Azalpena
NumTrees	uint8	Algoritmoaren garapenerako erabilitako zuhaitz kopurua
nMinLeafSize	uint8	Zuhaitz bakoitzeko erabilitako nodo kopurua
UMS	array<double>	$k=10$ partiziodun balioztatze gurutzatu bakoitzeko UMS balioak
MulAcc	array<double>	$k=10$ partiziodun balioztatze gurutzatu bakoitzeko MulAcc balioak

Taula 18. “Results.mat” fitxategian biltegituriko emaitzen egitura pertzentilak atziturik

Izena	Datu mota	Azalpena
NumTrees	uint8	Algoritmoaren garapenerako erabilitako zuhaitz kopurua
nMinLeafSize	uint8	Zuhaitz bakoitzeko erabilitako nodo kopurua
UMS	array<double>	$k=10$ partiziodun balioztatze gurutzatu bakoitzeko UMS balioak
MulAcc	array<double>	$k=10$ partiziodun balioztatze gurutzatu bakoitzeko MulAcc balioak
prctile_UMS	array<double>	$k=10$ partiziodun balioztatze gurutzatu bakoitzeko UMS balioei dagokien pertzentilak: P_{25} , P_{50} eta P_{75}
prctile_MulAcc	array<double>	$k=10$ partiziodun balioztatze gurutzatu bakoitzeko MulAcc balioei dagokien pertzentilak: P_{25} , P_{50} eta P_{75}

Deskribaturiko soluzioaren deskribapena laburbiltzen duen bloke diagrama Irudia 27an eskuragarri dago.

8.3.6. Emaitzen laburpena

Azkeneko atal honetan, proiektuan lortutako emaitzen laburpena aurkezten da. Alde batetik, emaitzak lortzeko prozedura Irudia 27an adierazi da eskematikoki, proiektuaren garapenerako buruturiko fase ezberdinak bistaratzen direlarik. Bertan, 4 azpi-atal bereiz daitezke: lehenengo atalak, Osakidetza Osasun Erakundeak lagaturiko datu-basean oinarriturik, segmentatutako eta anotatutako datu-basea garaturiko GUI interfazearen bitartez lortu dela aditzera ematen du. Bigarren atalak, algoritmoaren garapena hasi aurretiko fasea aurkezten du; horretarako, lehendabizi algoritmoaren garapenerako beharrezkoa den informazio esanguratsuz osaturiko datu-basea garatu da, bilduriko informazioa ezaugarritzen duten ezaugarrien erauzketa bigarren fasea izan da, eta ahalik eta algoritmorik errentagarriena lortzeko balioztatze gurutzatua aplikatu dela erakusten du. Hirugarren azpi-atalean bai, algoritmoaren garapenerako jarraituriko prozedura aurkezten da, azkenik, laugarren azpi-atalean, lorturiko algoritmo ezberdinen ebaluaketarako eta eraginkorrenaren hautaketarako erabilitako metrikak azaltzen direlarik. Bestetik, RF sailkatzailearen hainbat parametro aske doitu dira, froga guztien emaitza zehatzak I Eranskinean aurkezten direlarik. Jarraian emaitza guzti horien baliorik esanguratsuenak aurkeztu eta testuinguruan jarriko dira.

Proiektu konkretu honetan, lan egin den datu-baseak aurkezten duen erritmo klaseen desoreka dela medio, UMS balioetan oinarritu gara bereziki, hauek aurkezten duten sentikortasun orokorraren balioaren kalkulurako, erritmo klase bakoitzari pisu berdina eskaintzen zaio eta; hots, desoreka kontuan hartzen du. MulAcc balioak, kontrara, ez du klase despropozioa aintzat hartzen, lagin

bakoitzari pisu berdina emanaz. Badakigunez, gainera, 4 segundotako eta 8 segundotako seinale segmentuetan funtsaturiko sailkatzaile ezberdinak aztertu dira, horrela, bakoitzak bere aldetik aurkezten duen errendimendu maila ebaluatu ahal izateko.

Emaitzetan, orokorrean, gutxienezko behaketa kopurua handitu ahala, UMS balioek ere gora egin dutela, zehaztasun mailak, kontrara, behera egin duelarik ikusi da. Zuhaitz kopurua handitu ahala, ordea, UMS balioek, baita MulAcc balioek ere, biek gora egin dute. Gainera, sistemak aurkezten duen zehaztasun maila adierazten duen MulAcc balioen arteko aldeak oso txikiak izan direla ere azpimarratzekoa da; hau da, nahiz eta gutxienezko behaketa kopurua handitu ahala, zehaztasun mailak behera egin, hauen arteko aldea, bai 4 segmentuetan oinarritutako sailkatzaileetan zein 8 segmentuetan oinarritutakoetan ere, oso txikia izan da, %0.5 inguru. UMS balioen aldea, kontrara, handiagoa izan da: lehenengo kasuko algoritmo-sailkatzaileetan ($t=4$ seg segmentuak), aldea %4koa izan da gutxi gorabehera, bigarren kasuko sailkatzaileek ($t=8$ seg segmentuak), aldiz, %7rako jauzia jasanez. Aipatutako honek, datu-baseak aurkezten duen erritmo klase desorekaz gain, UMS sentikortasun maila adierazten duten emaitzak nabarmentzea eragin du.

Zortzi segundotako segmentuetan oinarrituta eraikitako sailkatzaileen kasuan, UMS balio onenak gutxienezko 15 eta 20 behaketekin buruturikoak izan dira, zuhaitzen arabera igoerak bilakaera argiago bat aurkezten dutelarik. Horrez gain, emaitzarik onenak 100 zuhaitzetatik gora erabiliz lortu dira, ordura arteko zuhaitz kopuruaren emaitzek apurka-apurka gora egiten dutelarik. Algoritmoaren zehaztasun mailari dagokionez ere, 100 zuhaitzetatik gora emaitza hobekak lortu dira, gutxienezko behaketa kopurua zenbat eta txikiagoa izan emaitza hobekak lortuz; hala ere, eta aurretik esan denez, aldeak hain dira txikiak, ezen bereziki garrantzia UMS balioei eman zaien, MulAcc balioak arbuizatu. Hori horrela, $t=8$ segundotako segmentuetan oinarrituriko sailkatzaileen artean eraginkorra eta zehaztasunik altuenekoa 175 zuhaitzekin eta gutxienezko 20 behaketekin eraikitakoa izan da.

Lau segundotako segmentuetan oinarrituta eraikitako sailkatzaileen kasuan, ordea, UMS sentikortasun maila erakusten duen metrikak ez du bilakaera handirik aurkeztu. Emaitza guztien artean, kaskarrenak gutxienezko behaketa kopurua 15 burutu dituen sailkatzaileena izan da, emaitzen artean, egonkortasunik ez, eta desoreka handia erakutsiz. Gainerako kasuetan, emaitzen bilakaera eta izaera nahiko egonkorra izan da, ezberdintasun bakarra, behaketa kopuru minimoa erabiltzetik (1) behaketa kopuru maximoa erabiltzera (20) lorturiko UMS balioek salto txiki bat jasan dutelarik (%3.8ko jauzia). Gutxienezko behaketa kopuru txikiarekin (1,3 eta 5) lorturiko algoritmoen UMS balioek egonkortasun eta izaera oso antzekoak aurkeztu dituzte, haien arteko aldea %1ekoa besterik izan gabe. Zehaztasuna adierazten duten MulAcc balioek, kontrara, zuhaitzen arabera bilakaera argiago bat aurkeztu dute, kasu honetan, 75-100 zuhaitz bitartean erabiltzetik aurrera egoera egonkortzen delarik; balio hauetatik aurrera, emaitzak oso antzeko mantendu dira. Hori dela eta, hemen ere, UMS balioari garrantzi handiagoa eman, eta beraz, 20 behaketekin eta zehaztasuna egonkortzen hasten den 75 zuhaitzekin (zuhaitz kopuru gehiagorekin prozesaketa denbora handiagoa behar da eta emaitzak oso antzekoak dira) eraikitako sailkatzailea hautatzea erabaki da.

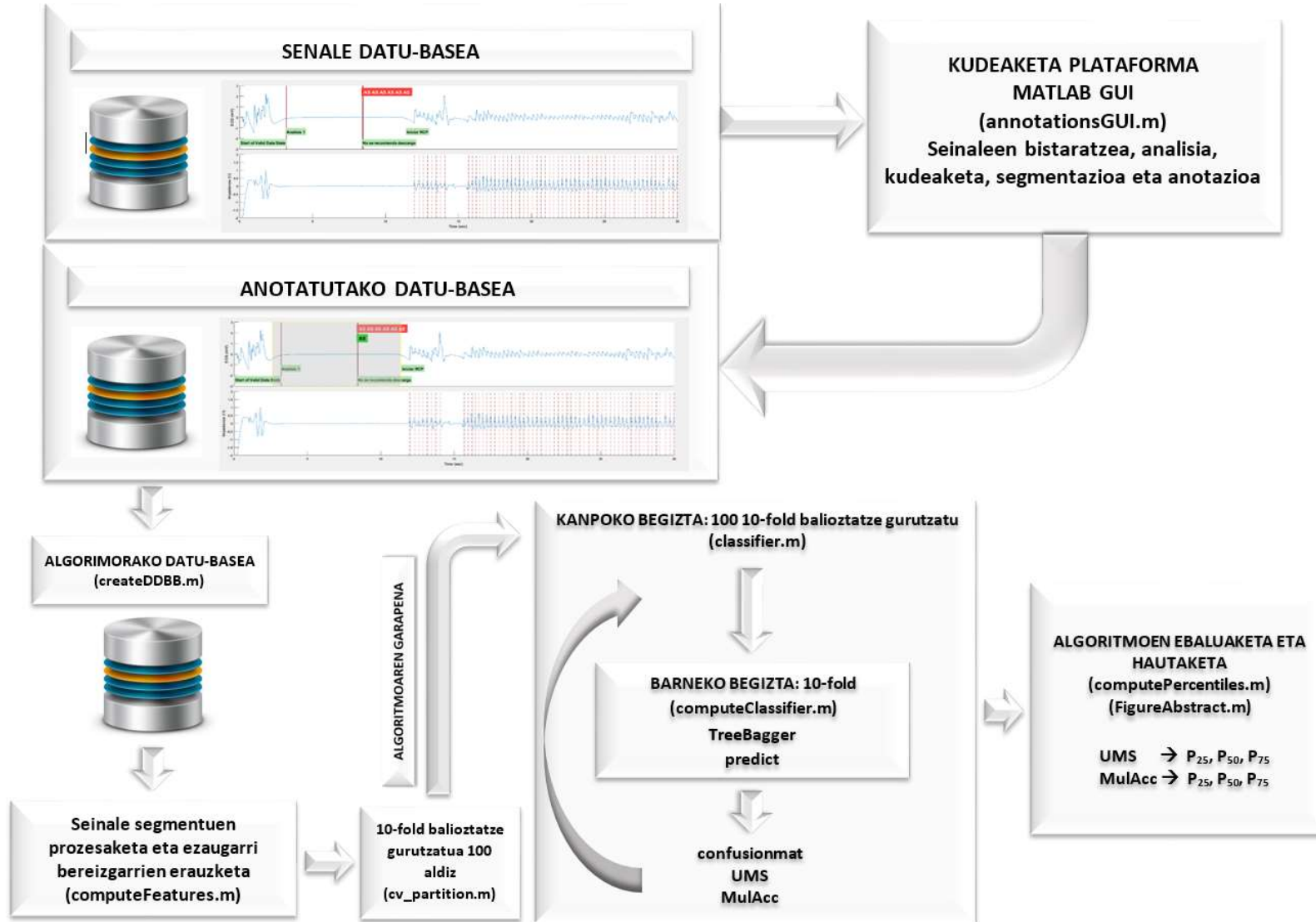
Honekin amaitzeko, 4 segundotako segmentuen sailkatzaileen emaitzak 8 segundotako segmentuetan oinarritutakoena baino kaskarragoak direla ondorioztatu da, hauen arteko UMS

sentikortasun aldea %5.8 puntutakoa eta sistemaren zehaztasun maila %0.6koa inguru izan dela baieztatuz. Jarraian, gailendutako sailkatzaileen emaitzak laburbiltzen dituen taula eskuragarri dago:

Taula 19. Gailendutako sailkatzaileen metriken emaitzak

METRIKA	T = 4seg		T = 8seg	
	UMS (%)	MulAcc (%)	UMS (%)	MulAcc (%)
MIN	80.10	96.45	85.83	96.50
MAX	84.37	96.84	90.01	96.87
P ₂₅	81.48	96.61	87.28	96.62
P ₅₀	81.65	96.66	88.44	96.69
P ₇₅	82.87	96.71	88.60	96.73
OUTLIER_1	-	-	84.61	-
OUTLIER_2	-	-	84.57	-
OUTLIER_3	-	-	84.47	-

Taulako emaitzei so eginez, orokorrean, edozein bi sailkatzaileekin lorturiko emaitzek errendimendu eta eraginkortasun altuak aurkezten dituztela agerikoa da; hala ere, 8 segundotako segmentuetan oinarritutakoen emaitzak hobekak dira. Kasu honetan, aldiz, askotan aipatu dugun datu-baseak aurkezten duen erritmo klaseen desoreka dela medio, MulAcc metrikaren emaitzak onak izatea gauza erraza da. Horregatik, etorkizun hurbil batean, beste algoritmo batzuekin zein proiektu honetan aztertzen den algoritmo egoera berdinekin ere lan egiteko, datu-base orekatuago eta osatuago batekin lan egitea gomendatuko nuke. Horrez gain, entrenamendu eta testeatze prozedurak datu-base zabalago batekin burutzeak, onura argiak ekarriko lituzke. Dena den, proiektu honetan buruturiko RF algoritmoarekin buruturiko konbinazio desberdinetako sailkatzaileen azterketak, RF algoritmoa berpizte bihotz erritmoen sailkapen automatikorako oso aproposa dela ondoriozta dezakegu, honek, aldi berean, eskaintzen dituen eraginkortasun eta erabilerraztasunari probetxua ateraz.



Irudia 27. Soluzioaren deskribapena laburbiltzen duen proiektuaren eskema orokorra

9. Metodologia

Metodologia atal honetan, proiektuaren garapenerako baliabide ezberdinak, baita burutuko diren lan-paketeen antolamendua deskribatuko dira.

9.1. Erabilitako baliabideak

Taula 20ak proiektua aurrera eramateko nahitaezkoak izan diren hardware eta software materialak laburbiltzen ditu:

Taula 20. Proiektuan zehar erabilitako baliabideak

ID	Materiala
MA1	Mahai gaineko DELL ordenagailua
MA2	Windows 7 sistema eragilea
MA3	Microsoft office word 2018
MA4	Matlab-eko R2017a lizentzia

9.2. Lan-paketeen deskribapena

Atal honetan, proiektuaren arrakasta bermatzeko proiektuan parte-hartzen duen lan-taldea, taldeko partaide bakoitzak jorratzen dituen ataza eta funtzioak definitu, baita lan-paketeak zeintzuk diren ere zehaztuko dira.

9.2.1. Lan-taldea

Taula 21ak lan-taldeko partaideak definitu, eta bakoitzak bete beharreko zereginak eta erantzukizunak zehazten ditu:

Taula 21. Proiektuan parte hartzen duen lan-taldearen deskribapena

Kodea	Erantzukizuna	Izen-Abizenak	Zeregina
K1	Proiektuaren zuzendaria	Unai Irusta Zarandona	Proiektua planteatzeaz, jarraitu beharreko pausuak adierazteaz eta dokumentua zuzentzeaz eta gainbegiratzeaz arduratzen da
K2	Junior ingeniaria	Munia Pitarke Ordoñez	Proiektuaren egilea. Proiektua garatzeaz eta idazteaz arduratzen da

9.2.2. Lan-paketeak

Jarraian, proiektua aurrera eramateko jarraitutako eta garatutako planifikazioaren atal eta fase ezberdinak definituko dira, eginkizunak eta hauen iraupenak ere zehaztuz. Proiektuak lau fase (lan-pakete) bereizten ditu: proiektuaren kudeaketa lehenengo fasea izango da, bigarren fasea proiektuaren prestakuntza, hirugarren fasea proiektuaren garapena eta azkeneko laugarren fasea proiektuaren aurkezpena. Proiektuaren lehenengo fasea labur deskribatzen duen lan-paketea:

Taula 22. Proiektuaren garapenerako lehenengo lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da

LP1	Hasiera	Bukaera	Iraupena (Egun)
PROIEKTUAREN KUDEAKETA	2018-02-19	2018-10-08	151
LP1.1: Lanaren kudeaketa, jarraipena eta ikuskapena: Proiektuaren arrakasta lortzeko nahitaezkoak diren koordinazio eta gainbegiratze lanak.	2018-02-19	2018-10-08	151

Bigarren fasean, proiektuaren helburu den berpizte bihotz erritmoen sailkatzaile automatikoa eraikitzeko ikasketa automatikoan oinarritutako algoritmoaren garapenerako beharrezkoa eta lagungarria den gutxienezko ezagutza eta prestakuntza deskribatzen da. Proiektuaren bigarren fasea laburbiltzen duen lan-paketea:

Taula 23. Proiektuaren garapenerako bigarren lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da

LP2	Hasiera	Bukaera	Iraupena (Egun)
PROIEKTUAREN PRESTAKUNTZA	2018-02-19	2018-03-23	25
LP2.1: Aurretiko formakuntza: Proiektua garatu eta jorratu aurretiko beharrezkoa eta lagungarria izan den oinarritzko kontzeptuen ikasketa: GULaren erabilpena sakondu.	2018-02-19	2018-03-02	10
LP2.2: Gaiaren inguruko ezagutza sakondu: Kardiologia eremuan oinarriturik bihotz-biriketako gelditzearen aurrean berpiztearen eta Machine Learning inguruan beharrezkoa den gutxienezko jakituria eta erabiliko diren baliabideen inguruko ezagutza lortu: artikuluen irakurketa... Honek, proiektua garatzeko orduan erraztasun asko ekarriko ditu.	2018-03-05	2018-03-16	10
LP2.3: Proiektuaren zehaztapena: Atal honen xede nagusia proiektuaren helburu nagusiak eta irismena zeintzuk diren	2018-03-19	2018-03-23	5

zehaztea eta definitzea izango da; hots, plangintzaren aurretiko zehaztapena.			
---	--	--	--

Hirugarren fasean, berpizte bihotz erritmoen sailkapen automatiko eraginkor baten garapenerako buruturiko ataza ezberdinak laburbiltzen dira. Proiektuaren hirugarren fasea labur deskribatzen duen lan-paketea, beraz:

Taula 24. Proiektuaren garapenerako hirugarren lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da

LP3	Hasiera	Bukaera	Iraupena (Egun)
PROIEKTUAREN GARAPENA	2018-03-26	2018-08-09	86
<p>LP3.1: GUIaren garapena:</p> <p>E3.1.1 Eginkizuna: BERPizte bihotz erritmoen sailkatzailea sortzeko ezinbestekoa den datu-basea osatzen duten seinaleen bistaraketa, analisia, anotazioa eta segmentazioa ahalbidetzen duen plataformaren garapena.</p>	2018-03-26	2018-04-27	23
<p>LP3.2: Datu-basearen anotazioa eta segmentazioa:</p> <p>E3.2.1 Eginkizuna: Seinaleen banakako analisi tarte bakoitzeko analisia, anotazio bateratzea eta segmentazioa.</p>	2018-04-30	2018-05-28	20
<p>LP3.3: Algoritmorako anotatutako eta segmentatutako datu-basearen sorrera:</p> <p>E3.3.1 Eginkizuna: Algoritmoaren garapenerako beharrezko laginak biltzen dituen datu-basea sortu.</p>	2018-05-29	2018-05-29	1
<p>LP3.4: Algoritmoaren garapena:</p> <p>E3.4.1 Eginkizuna: Seinale segmentuen prozesaketa.</p> <p>E3.4.2 Eginkizuna: Seinale segmentuak deskribatzen dituzten ezaugarri bereizgarrien erauzketa.</p> <p>E3.4.3 Eginkizuna: Algoritmo eraginkor baten garapenerako datu-baseko laginetan balioztatze gurutzatua aplikatzea.</p> <p>E3.4.4 Eginkizuna: Machine Learning ikasketa automatikoan funtsaturiko Random</p>	2018-05-30	2018-06-05	5
	2018-06-06	2018-06-12	5
	2018-06-13	2018-06-19	5

Forest algoritmoaren exekuzioa. Entrenamendu eta balioztatze faseak. E3.4.5 Eginkizuna: Confusion Matrix matrizearen eta ebaluazio metriken kalkulua: UMS eta MulAcc	2018-06-20	2018-07-02	9
LP3.5: Emaizten azterketa eta algoritmoaren hautaketa: E3.5.1 Eginkizuna: Metriken araberako emaitzen bistaratzea. E3.5.2 Eginkizuna: Emaizten konparaketa eta analisia. E3.5.3 Eginkizuna: Emaizten araberako algoritmo eraginkorrenaren aukeraketa	2018-07-03	2018-07-05	3
	2018-07-06	2018-08-09	15

Proiektuaren xedea lorturik, azkeneko laugarren fasea aurkezten duen laugarren lan-paketea:

Taula 25. Proiektuaren garapenerako hirugarren lan-paketeari dagokion proiektuaren kudeaketa deskribatzen da

LP4	Hasiera	Bukaera	Iraupena (Egun)
PROIEKTUAREN AURKEZPENA	2018-08-10	2018-10-08	40
LP4.1: Proiektuaren dokumentazioa: Lana era argiengan deskribatzen duen agiria: proiektuaren helburu nagusiak, norainokoa/irismena, gizartearentzako biltzen dituen onuren multzoa, gaur egungo egoera, proiektua garatzeko metodologia, plangintza eta abar.	2018-08-10	2018-09-24	30
LP4.2: Proiektuaren ahozko aurkezpena: Garatutako proiektua defendatzeko derrigorrezko epaimahaiaren aurreko aurkezpena.	2018-09-25	2018-10-08	10

Jarraian, proiektuaren programazioa denbora unitateetan laburbiltzen duen taula aurkezten da:

Taula 26. Lan-unitateen deskribapena

UNITATEA	DESKRIBAPENA
PROIEKTUA	30
ASTEAK	5
EGUNA	5

9.2.3. Mugarriak eta entregagaiak

Lehenengo taula honek, proiektua arrakastatsu garatzeko eta bukatzeko mugarriak aurkezten ditu:

Taula 27. Proiektuaren mugarriak

Kodea	Mugarriaren azalpena	Data
M1	Proiektuaren hasiera	2018-02-19
M2	Aurretiko formakuntzaren amaiera	2018-03-02
M3	Proiektuaren definizioa eta izaera	2018-03-23
M4	GUI kudeaketa plataformaren amaiera	2018-04-27
M5	Datuen anotazioen eta segmentazioen amaiera	2018-05-28
M6	Algoritmorako datu-basearen sorreraren amaiera	2018-05-07
M7	Algoritmoaren garapenaren amaiera	2018-07-02
M8	Emaitzen amaiera	2018-07-05
M9	Algoritmoaren hautaketaren amaiera	2018-08-09
M10	Proiektuaren garapenaren amaiera	2018-08-09
M11	Proiektuaren dokumentuaren garapenaren amaiera	2018-09-24
M12	Proiektuaren aurkezpenaren garapenaren amaiera	2018-10-08
M13	Proiektuaren amaiera	2018-10-08

Proiektua garatu ahala lortzen diren nahitaezko entregagaiak zeintzuk diren deskribatuko dira atal honetako bigarren taulan:

Taula 28. Proiektuaren entregagaiak

Kodea	Entregagaien azalpena	Data
E1	Anotatutako eta segmentatutako datu-basea	2018-05-29
E2	Algoritmoak eta hauen emaitzak eta ondorioak	2018-08-09
E3	Proiektuaren memoria dokumentua	2018-09-24
E4	Aurkezpena	2018-10-08

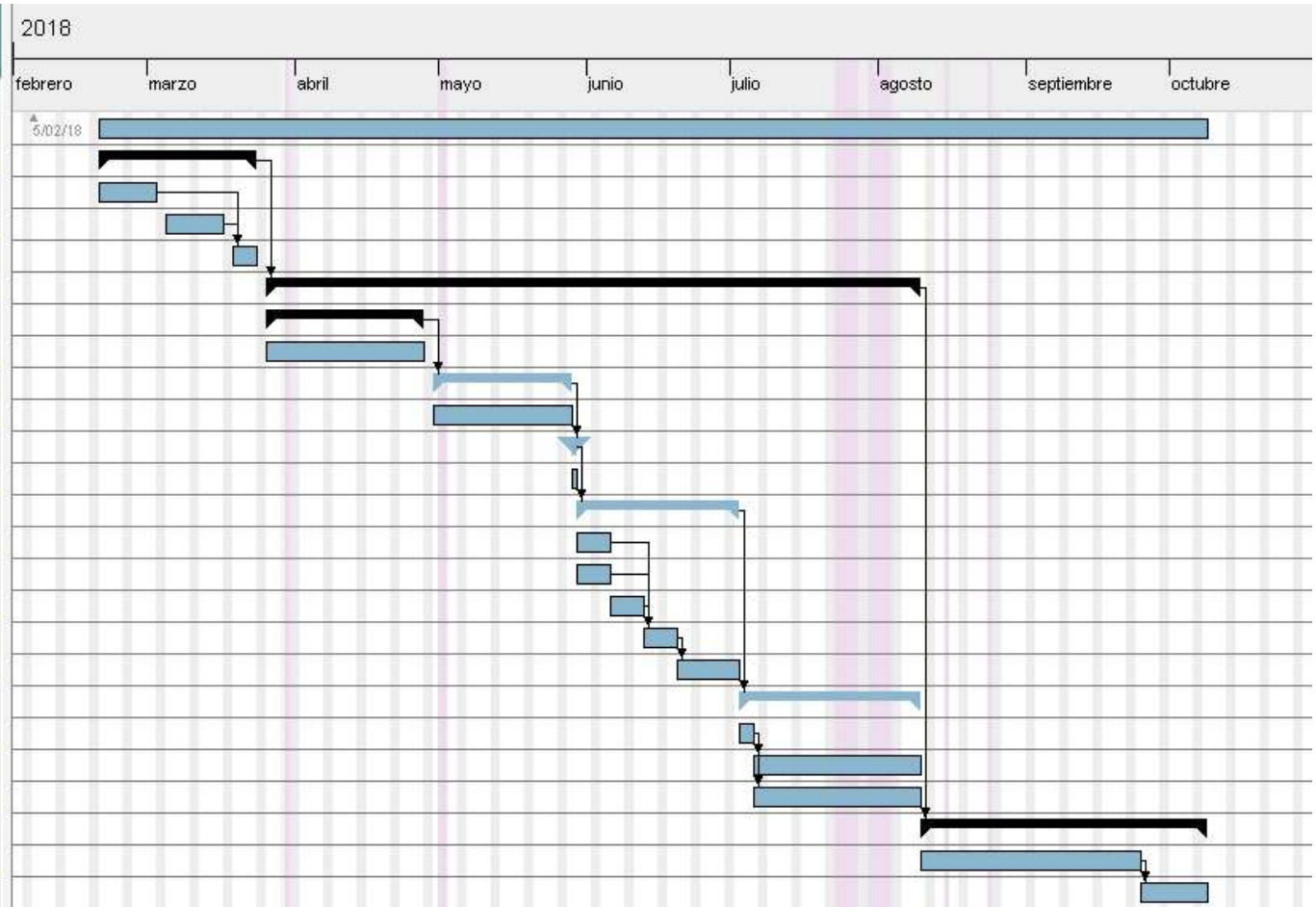
9.2. Gantt diagrama

Gantt diagramak irakurtzeko eta ulertzeko erraztasun handia izateagatik, proiektu gehienetan erabilia izaten den erreminta dugu. Honekin, proiektuaren aurrerapenaren eta garapenaren adierazpide grafikoa era erraz, simple eta argi batean aurkezten da, langileen arteko komunikazioa asko errazteko eta hobetzeko baliabide eraginkorra izanik.

Jarraian, deskribaturiko eta proiektuan zehar jarraitutako planifikazioa era oso argi batean aurkezten duen Gantt diagrama aztertzeko aukera dago.



Nombre	Fecha de inicio	Fecha de fin
• LP1. Proiektuaren kudeaketa	19/02/18	8/10/18
▣ • LP2. Proiektuaren prestakunt...	19/02/18	23/03/18
• LP2.1	19/02/18	2/03/18
• LP2.2	5/03/18	16/03/18
• LP2.3	19/03/18	23/03/18
▣ • LP3. Proiektuaren garapena	26/03/18	9/08/18
▣ • LP3.1	26/03/18	27/04/18
• E3.1.1	26/03/18	27/04/18
▣ • LP3.2	30/04/18	28/05/18
• E3.2.1	30/04/18	28/05/18
▣ • LP3.3	29/05/18	29/05/18
• E3.3.1	29/05/18	29/05/18
▣ • LP3.4	30/05/18	2/07/18
• E3.4.1	30/05/18	5/06/18
• E3.4.2	30/05/18	5/06/18
• E3.4.3	6/06/18	12/06/18
• E3.4.4	13/06/18	19/06/18
• E3.4.5	20/06/18	2/07/18
▣ • LP3.5	3/07/18	9/08/18
• E3.5.1	3/07/18	5/07/18
• E3.5.2	6/07/18	9/08/18
• E3.5.3	6/07/18	9/08/18
▣ • LP4. Proiektuaren aurkezpena	10/08/18	8/10/18
• LP4.1	10/08/18	24/09/18
• LP4.2	25/09/18	8/10/18



Irudia 29. Proiektuaren metodologiaren Gantt diagrama

10. Alderdi ekonomikoak

Atal honetan, proiektuaren garapenerako ezinbestekoak diren baliabideen aurrekontuak barnean hartzen dituen kontu-sailak xehetasunez aurkezten da, ondoren, aurrekontuaren kostu totala zenbatekoa den kalkulaturik:

10.1. Aurrekontuaren kontu-sailak

BARNE-ORDUAK

Barne-orduek proiektua aurrera eramateko eta burutzeko parte-hartzaileen orduak definitzen dituzte, hauen kostu totala ere adieraziz. Proiektu honetan proiektu zuzendari batek eta junior ingeniari batek parte hartu dute beraien kostu unitarioak hurrengoak izanik:

Taula 29. Proiektuan parte hartzen dutenen eskulanaren prezio unitarioa

Kodea	Kargua	Kostua(€/h)
K1	Proiektuaren zuzendaria	60
K2	Junior ingeniaria	30

Hurrengo taulan, partaideek ataza bakoitzean jardundako ordu kopurua eta hauen kostuen jorraketa laburbilduko dira; hots, proiektu honi egokitzen zaion giza baliabideen kostuen jorraketa.

Taula 30. Aurrekontuko giza baliabideen kostuen hedapena eta jorraketa

Ataza	Langilea	Kostua(€/h)	Lan orduak	Kostu totala (€)
LP1.1	K1	60	15	900
	K2	30	20	600
LP2.1	K1	60	0	0
	K2	30	20	600
LP2.2	K1	60	2	120
	K2	30	50	1500
LP2.3	K1	60	10	600
	K2	30	10	300
LP3.1	K1	60	5	300
	K2	30	115	3450
LP3.2	K1	60	90	5400
	K2	30	100	3000
LP3.3	K1	60	0	0
	K2	30	2	60
LP3.4	K1	60	5	300
	K2	30	96	2880
LP3.5	K1	60	10	600

	K2	30	90	2700
LP4.1	K1	60	25	1500
	K2	30	150	4500
LP4.2	K1	60	10	600
	K2	30	50	1500

Datu hauekin guztiekin aurrekontu osoaren barnean hartzen den barne-orduei dagokien laburpen taula bistaratzen da jarraian:

Taula 31. Barne-orduen laburpen taula

Behargina	Kostua(€/h)	Lan orduak	Kostu totala (€)
K1 Proiektuaren zuzendaria	60	172	10320
K2 Junior ingeniaria	30	703	21090
Guztira	-	-	31410 €

AMORTIZAZIOAK/INBERTSIOAK

Amortizazioak proiektua gauzatzeko eta garatzeko erabilitako aktibo finkoak definitzen ditu, hau da, erabilitako baliabide materialak zein materialak ez direnak definitzen ditu. Beheko taulak, proiektu honetan erabilitako aktibo finkoen amortizazioen kostu totala laburbiltzen du:

Taula 32. Amortizazioen laburpen taula

Kontzeptua	Hasierako prezioa (€)	Bizitza erabilgarria (hilabete)	Erabiltze denbora (hilabete)	Kostu totala (€)
Mahai gaineko DELL ordenagailua	600	48	7.5	93.75
Windows 7 sistema eragilea	500	36	7.5	104.2
Microsoft office word 2018	150	24	2	12.5
Matlab-eko R2017a lizentzia	2000	24	6.5	541.7
Guztira	-	-	-	752.12 €

GASTUAK

Proiekturako erabiltzen diren kontzeptuak eta behin erabilia berriro erabili ezin daitezkeenak; hots, gastuak kontu sailak barnean kontzeptu suntsigarriak hartzen ditu. Jarraian bistaratzen den 17.Taulak hauen kostu totalen laburpena erakusten du:

Taula 33. Gastuen laburpen taula

Kontzeptua	Prezioa (€)	Unitate kopurua	Kostu totala (€)
Bulegoko materiala	20	-	20
Argia	40	-	40
Kableak	2	-	2
16 Gbyteko Lexar prendrive-a	20	1	20
Guztira	-		82 €

10.2. Aurrekontu osoaren laburpena

Jarraian, aipatutako kontu-sailak barnean hartuz, proiektu honi egokitzen zaion aurrekontu osoaren laburpen taula bistaratuko da:

Taula 34. Aurrekontu osoaren laburpen taula

Kontzeptua	Kostua (€)
Barne-orduak	31410
Amortizazioak/Inbertsioak	752.12
Gastuak	82
TOTALA (BEZ gabe)	32244.12 €

Proiektuaren garapenean zeharreko parte hartzaileen barne-orduen, amortizazioen, inbertsioen eta gastuen kostu totalak kontuan harturik, lortutako aurrekontuaren kostua **hogeita hamabi mila berrehun eta berrogeita lau koma hamabi** eurokoa da.

11. Ondorioak

Proiektuan zehar askotan aipatu denez, proiektu honek bilatzen duen xede nagusia, berpizte bihotz erritmoak identifikatzen eta sailkatzen dituen algoritmo automatiko baten garapenean oinarritzen da. Mundu mailan sistema kardiobaskularrarekin eta arnas sistemarekin hertsiki loturiko medikuntza esparruan lan egiten duten talde ikertzaileen sustapenari esker, azken urteotako etengabeko garapena ekarri duten bihotz arritmiak bereizteko gaitasuna aurkezten duten EKG seinaleen algoritmo-sailkatzaile ugari sortu dira. Proiektu honek eskaintzen dituen ekarpen zientifikoek eta teknikoek, beraz, ikerkuntza elkarrekin, bai eta egunerokoan aurkeztu daitezkeen larrialdi zein etengabeko monitorizazio egoeretan onuragarriak, baliagarriak eta mesedegarriak izatea espero da.

Proiektuak eskaintzen duen ekarpenetako bat, biomedikuntza alorrean berebiziko garrantzia duen bihotz berpizte seinaleen aurretiko analisi, anotazio, identifikazio eta segmentazio doituak eta zehatzak egitea ahalbidetzen duen tresna baten garapena da. Gainera, eraikitako interfaze grafiko honi esker, seinaleen azterketan zehar ezinbestekoak diren buruturiko anotazio eta segmentazioak biltzen dituen datu-base baten garapenerako gaitasuna ematen da, aurrerago proiektu honetan burutu izan den algoritmoaren garapenerako, zein etorkizunean sortu daitezkeen algoritmoen oinarri gisa erabiltzeko gaitasuna eskaintzen duelarik.

Hala ere, proiektuak eskaintzen duen ekarpenik esanguratsuena, berpizte bihotz erritmoen sailkatzaile automatikoaren garapena da. Kasu honetan, MLak eskuragarri dituen algoritmo ezberdinen arteko aldeak aztertu beharrean, Random Forest algoritmoa bera bakarrik aplikatuta lor daitezkeen parametro konbinazio ezberdinen (zuhaitz kopuruaren eta zuhaitz nodo bakoitzaren gutxienezko behaketa kopuruaren araberrako konbinazioak) araberrako algoritmo-sailkatzaile ezberdinen arteko aldeak dira aztertu direnak. Gainera, egunerarte garatu izan diren beste algoritmo batzuekin alderaturik, hauek sailkatzailean konfiantza tarte nabarmenak sortzen dituzten 100 aldiz errepikaturiko balioztatze gurutzatua darabilte, algoritmoa are sendoagoa izatea lortzen delarik. Horrela, RF algoritmoak aurkezten dituen malgutasuna, erraztasuna, abiadura eta doitasuna baliatuz, algoritmo-sailkatzaile eraginkorra lortzea posiblea izan da.

Lorturiko emaitzarik eraginkorrenak, 175 zuhaitzekin eta gutxienezko 20 behaketekin buruturiko algoritmo-sailkatzailea izan da, erdietsitako zehaztasun orokorra eta erritmo klaseen proportzioa errespetatzen duen sentikortasuna %96.5 eta %88koa izanik hurrenez hurren. Emaitzak etorkizun hurbilerako onuragarriak direla agerikoa da, baita hobetu daitezkeela ere, entrenamendu eta testeatze prozeduretarako datu-base zabalago bat, zein erritmo klaseen arteko proportzio orekatuago bat lortuz adibidez. Etorkizun hurbil batean, beraz, aipaturiko datu-base zabalago eta orekatuago batekin lan egitea ondo legoke. Horrez gain, sorturiko interfaze grafikoan zein datu-basean funtsaturik, ML algoritmo ezberdinak aplikatzeko eta lorturiko emaitzak analizatzeko ere aukera ona litzateke.

Laburbilduz, beraz, honekin guztiarekin berpizte bihotz erritmoen sailkapen automatikorako garaturiko ikasketa automatikoan oinarrituriko algoritmoari ahalik eta onura eta potentzialik handiena ateratzea espero da. Hau da, herrialde garatuenean lehen heriotza-arrazoia bihotz-heriotza izanik, bihotz-biriketako geldiunean dagoen gizakiaren bizi-aukera handitzeko algoritmo, teknika eta metodo berriak bilatzen dituzten ikerketen sustapenerako erraztasunak eta laguntza ematea espero da. Horrez gain, etengabeko kontrolean dauden pazienteen monitorizazioetan, baita egoera kritikoan aurkitzen

diren biktimen egoeren aurrean erritmoen bilakaera aurrerako erraztasunak eskaintzea zein, terapia egokien hautapena eta erabilpena berehalakoa eta egokia izateko aukera ematea ere espero da.

Erreferentziak

- [1] Gómez, N., Cañizares, M., González, R.I. eta Rivero, M.M., Un algoritmo para la clasificación del ritmo cardiaco. IFMBE Proc. 2004; 5(1): 1359-1362.
- [2] Rad, A.B., Eftestøl, T., Engan, K., Irusta, U., Kvaloy, J.T., Kramer-Johansen, J. et al, ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. IEEE Trans Biomed Eng., 2017; 64: 2411–2418.
- [3] Rad, A.B., Eftestøl, T., Irusta, U., Kvaloy, J.T., Wik, L., Kramer-Johansen, J., K. Katsaggelos, A. et Engan, K., An automatic system for the comprehensive retrospective analysis of cardiac rhythms in resuscitation episodes. Resuscitation, 2018; 122: 6-12.
- [4] Ramírez López, L., Model of dynamic classification of arrhythmias cardiac by means of learning of machine with interface to user. Ciencia e Ingeniería Neogranadina, 2006; 16: 86-95.
- [5] Rad, A.B., Engan, K., K. Katsaggelos, A., Kvaloy, J.T., Wik, L., Kramer-Johansen, J., Irusta, U. et Eftestøl, T., Automatic cardiac rhythm interpretation during resuscitation. Resuscitation, 2016; 102: 44-50.
- [6] Alonso, E., Aramendi, E., Daya, M., Irusta, U., Chicote, B., K. Russell, J. et G. Tereshchenko, L., Circulation detection using the electrocardiogram and the thoracic impedance acquired by defibrillation pads. Resuscitation, 2016; 99; 56-62.
- [7] Soar, J., Nolan, J.P., Böttiger, B.W., Perkins, G.D., Lott, C., Carli, P., Pellis, T., Sandroni, C., Skrifvars, M.B., Smith, G.B., Sunde, K. et Deakin, C.D., European Resuscitation Council Guidelines for Resuscitation 2015Section 3. Adult advanced life support. Resuscitation, 2015; 95: 100-147.
- [8] Stiell, I.G., MD, MSc, Wells, G.A., PhD, Herbert, P.C., MD, MHSc, Laupacis, A., MD, MSc, Weitzman, B.N., MD, Association of Drug Therapy with Survival in Cardiac Arrest: Limited Role of Advanced Cardiac Life Support Drugs. Academic Emergency Medicine, 2015; 2(4): 264-273.
- [9] Vincent, R., Drugs in modern resuscitation. British Journal of Anaesthesia, 1997;79:188-197.
- [10] Bravo, D. et Godoy, J., Drogas y volumen en la reanimación cardiopulmonar. Rev. Chil Anest, 2012; 41: 36-41.
- [11] Bolón-Canedo, V., Remeseiro, B., Alonso-Betanzos, A. et Campilho, A., Machine Learning for medical applications. ESANN, 2016; 225-234. Eskuragarri: <http://www.i6doc.com/en/>
- [12] Raval, D., Bhatt, D., Kumhar, M.K., Parikh, V. et Vyas, D., Medical Diagnosis System Using Machine Learning. IJCS, 2015-2016; 7(1): 177-182.
- [13] Vishwa, A., Lal, M.K., Dixit, S. et Vardwaj, P., Clasificación Of Arrhythmic ECG Data Using Machine Learning Techniques. Internationak Journal of Artificial Intelligence and Interactive Multimedia, 2011; 1(4): 68-71.
- [14] Paiva, D., Automatic Arrhythmia Classification: A Pattern Recognition Approach. Biomedical Engineering Thesis, 2014.

- [15] Wei-Yin Loh., Classification and regression trees. *WIREs Data Mining and knowledge Discovery*, 2011; 1: 14-23.
- [16] Cánovas-García, F., Alonso-Sarría, F. et Gomariz-Castillo, F., Modificación del algoritmo Random Forest para su empleo en clasificación de imágenes de teledetección. *XVII Congreso Nacional de Tecnologías de información Geográfica*, 2016: 359-368.
- [17] Raghav Aggiwal, Introduction to Random forest, 2017: Eskuragarri: <https://16.in/introduction-to-random-forest/> (Kontsulta: 2018 irailak 9).
- [18] Sabater Morales, Alberto, Algoritmos de aprendizaje automatizado para el meta compresor de datos Blosc, 2017.
- [19] Breiman, Leo, RANDOM FORESTS, 2001
- [20] Medina Merino, R.F. et Ñique Chacón, C.I., BOSQUES ALEATORIOS COMO EXTENSIÓN DE LOS ÁRBOLES DE CLASIFICACIÓN CON LOS PROGRAMAS R Y PYTHON. *INTERFASES*, 2017; 10: 165-189.
- [21] Breiman L. Bagging predictors. *Mach. Learning*, 1996; 24: 123-140.
- [22] Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X. et Lu, Z., MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Research*, 2007; 35: 339-344.
- [23] Li, T. et Zhou, M., ECG Classification Using Waveley Packet Entropy and Random Forests. *Entropy*, 2016; 18(285). Eskuragarri: www.mdpi.com/journal/entropy.
- [24] Bootstrap Sample: Definition, Example. Eskuragarri: <http://www.statisticshowto.com/bootstrap-sample/> (Kontsulta: 2018 irailak 9).
- [25] Leoca George, Doru, Estudio de simulación estocástica para el análisis de las propiedades de los estimadores Random Forests frente a árboles individuales.
- [26] Figuera, C., Irusta, U., Morgado, E., Aramendi, E., Ayala, U., Wik, L., Kramer-Johansen, J., Eftestøl, T. et Alonso-Atienza, F., Machine Learning Techniques for the Detection of Shockable Rhythms in Automated External Defibrillators. *PLOS ONE*, 2016; 11(7). DOI:10.1371/journal.pone.0159654.
- [27] Thakor NV, Zhu YS, Pan KY. Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Transactions on Biomedical Engineering*, 1990. September; 37(9): 837-43. doi: 10.1109/10.58594.
- [28] Arafat M, Chowdhury A, Hasan M. A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *Signal, Image and Video Processing*, 2011; 5: 1-10. doi: 10.1007/s11760-009-0136-1.
- [29] Amann A, Tratnig R, Unterkofler K., Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. *Biomed Eng Online*. 2005; 4:60.

- [30] Anas E, Lee S, Hasan M, Sequential algorithm for life threatening cardiac pathologies detection based on mean signal strength and EMD functions. *BioMedical Engineering OnLine*. 2010; 9(1):4-3 doi: 10.1186/1475-925X-9-43.
- [31] Jekova I, Krasteva V, Real time detection of ventricular fibrillation and tachycardia. *Physiological Measurement*. 2004; 25(5): 1167 doi: 10.1088/0967-3334/25/5/007.
- [32] Ayala U, Irusta U, Ruiz J, Eftestøl T, Kramer-Johansen J, Alonso-Atienza F, Alonso E, González-Otero D., A reliable method for rhythm analysis during cardiopulmonary resuscitation. *Biomed Res Int*. 2014; 2014: 872470.
- [33] Irusta U, Ruiz J, Aramendi E, Ruiz de Gauna S, Ayala U, Alonso E., A high-temporal resolution algorithm to discriminate shockable from nonshockable rhythms in adults and children. *Resuscitation*. 2012 Sep; 83(9): 1090-7.
- [34] - Kuo S, Dillman R. Computer detection of ventricular fibrillation In: *Computers in Cardiology*; 1978. p. 2747–2750.
- [35] Barro S, Ruiz R, Cabello D, Mira J., Algorithmic sequential decision making in the frequency domain for life threatening centricular arrhythmias and aimitative artifacts: a diagnostic system. *Journal of Biomedical Engineering*. 1989. July; 11(4): 320–8. doi: 10.1016/0141-5425(89)90067-8.
- [36] Li Y, Bisera J, Weil MH, Tang W., An Algorithm Used for Ventricular Fibrillation Detection Without Interrupting Chest Compression. *IEEE Transactions on Biomedical Engineering*. 2012. January; 59(1): 78–86. doi: 10.1109/TBME.2011.2118755.
- [37] Zhang XS, Zhu YS, Thakor NV, Wang ZZ., Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering*. 1999. May; 46(5): 548–55. doi: 10.1109/10.759055.
- [38] Jekova I. Shock advisory tool: Detection of life-threatening cardiac arrhythmias and shock success prediction by means of a common parameter set. *Biomedical Signal Processing and Control*. 2007; 2(1): 25–33. doi: 10.1016/j.bspc.2007.01.002.
- [39] Amann A, Tratnig R, Unterkofler K., Detecting ventricular fibrillation by time-delay methods. *IEEE Transactions on Biomedical Engineering*. 2007. January; 54(1): 174–177. doi: 10.1109/TBME.2006.880909.
- [40] Lake DE, Richman JS, Griffin MP, Moorman JR., Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol*. 2002. September; 283(3): R789–797. doi: 10.1152/ajpregu.00069.2002.
- [41] Eskuragarri: <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76> (Kontsulta: 2018 irailak 9).
- [42] Eskuragarri: <http://pitia.unileon.es/varp/sites/default/files/PublicationPDF/Clasificacion%20y%20reconocimiento%20de%20patrones.pdf> (Kontsulta: 2018 irailak 9).
- [43] Mar, T., Zaunseder, S., Martínez, J.P., Llamedo, M. eta Poll, R., Optimization of ECG Classification by Means of Feature Selection. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 2011; 58(8): 2168-2177.

[44] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. eta Duchesnay, E., SCIKIT-LEARN: MACHINE LEARNING IN PYTHON. Journal of Machine Learning Research, 2011; 12: 2825-2830.

[45] “Python orrialde nagusia.” Eskuragarri: <https://www.python.org/> (Kontsulta: 2018 irailak 6).

[46] Eskuragarri: <https://www.pythonforbeginners.com/learn-python/what-is-python/> (Kontsulta: 2018 irailak 6).

[47] Eskuragarri: https://www.codementor.io/innat_2k14/preferable-tools-for-machine-learning-python-matlab-r-jfozzpzhz_ (Kontsulta: 2018 irailak 6).

[48] Eskuragarri: <https://revistadigital.inesem.es/informatica-y-tics/python-r-analisis-datos/> (Kontsulta: 2018 irailak 6).

[49] Eskuragarri: <https://www.ecured.cu/MATLAB> (Kontsulta: 2018 irailak 6).

[50] Eskuragarri: <http://nereida.deioc.ull.es/~pcgull/ihiu01/cdrom/matlab/contenido/node2.html> (Kontsulta: 2018 irailak 6).

[51] MATLAB web orrialde ofiziala. Eskuragarri: <http://es.mathworks.com/products/matlab/> (Kontsulta: 2018 irailak 6).

[52] MATLAB lizentziaren prezioa. Eskuragarri: <http://es.mathworks.com/pricing-licensing/> (Kontsulta: 2018 irailak 6).

[53] GNU Octave. Eskuragarri: https://es.wikipedia.org/wiki/GNU_Octave (Kontsulta: 2018 irailak 6).

[54] Eskuragarri: <http://tododelenguajec.blogspot.com/2011/11/caracteristicas-del-lenguaje-c.html> (Kontsulta: 2018 irailak 6).

[55] Eskuragarri: <http://gigatecno.blogspot.com/2014/05/ventajas-y-desventajas-de-programar-en-c.html> (Kontsulta: 2018 irailak 6).

[56] Eskuragarri: <https://es.wikipedia.org/wiki/C> (Kontsulta: 2018 irailak 6).

[57] Díaz Gómez, F., Aprendizaje automático: métodos y aplicaciones. Eskuragarri: <http://www.eui.uva.es/doc/anuncios/curso.2013.2014/aprendizajeAutomatico.pdf> (Kontsulta: 2018 irailak 6).

[58] Gago Utrera, R., Uso de algoritmos de aprendizaje automático aplicados a bases de datos genéticos(2017).Eskuragarri:<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/65426/6/rgagoTFM0617memoria.pdf> (Kontsulta: 2018 irailak 6).

- [59] Machine Learning con Python. Eskuragarri: <https://relopezbriega.github.io/blog/2015/10/10/machine-learning-con-python/> (Kontsulta: 2018 irailak 6).
- [60] Eskuragarri: <http://www2.famaf.unc.edu.ar/institucional/biblioteca/trabajos/638/17067.pdf> (Kontsulta: 2018 irailak 6).
- [61] Sáenz Bajo, N. eta Ballesteros, M.A., Redes neuronales: concepto, aplicaciones y utilidad en medicina. Reflexiones en medicina de familia, 2002; 30(2): 119-120.
- [62] Ruiz, C.A. eta Basualdo, M.S., Redes Neuronales: Conceptos Básicos y Aplicaciones (2001).
- [63] Random Forest. Eskuragarri: <http://randomforest2013.blogspot.com/2013/05/randomforest-definicion-random-forests.html> (Kontsulta: 2018 irailak 6).
- [64] Fundamentos para la dirección de proyectos (Guía del PMBOK). Cuarta Edición. 2008 Project Management Institute.
- [65] Eskuragarri: <https://es.slideshare.net/draurcuyo81/caso-de-asistolia> (Kontsulta: 2018 irailak 6).
- [66] Eskuragarri: <http://www.my-ekg.com/trucos-consejos-ekg/ritmo-sinusal-diagnostico.html> (Kontsulta: 2018 irailak 6).
- [67] Eskuragarri: <https://www.practicalclinicalskills.com/ekg-reference-guide-details-es?lessonid=26> (Kontsulta: 2018 irailak 6).
- [68] Corsiglia, D. eta Fontana, J.J., Arritmias cardíacas. Eskuragarri: http://www.fac.org.ar/edicion/guias_rcp/Cap9.pdf (Kontsulta: 2018 irailak 6).
- [69] García Gutiérrez, A., Machine Learning en Bases de Datos de Lenguaje Natural (2016). Eskuragarri: https://repositorio.uam.es/bitstream/handle/10486/676778/Garcia_Gutierrez_Alvaro_tfg.pdf?sequence=1 (Kontsulta: 2018 irailak 6).

I ERANSKINA

Algoritmoen emaitzak

I ERANSKINA: Algoritmoen emaitzak

Dokumentuaren atal honetan, ML ikasketa automatikoak eskaintzen duen Random Forest algoritmoa eta honek eskuragarri dituen parametro ezberdinen arabera lortutako berpizte bihotz erritmoen sailkatzaile ezberdinen emaitza esanguratsuenak eta adierazgarrienak aurkeztuko dira; ildo honetatik, erdietsitako sailkatzaile guztien artean eraginkorrena eta errendimendurik altuena aurkezten duena ere aukeratuko da.

Badakigunez, Random Forest algoritmo sailkatzaileak eskuragarri dituen parametroen artean, geuk zuhaitz kopurua eta zuhaitz bakoitzak izan ditzakeen nodo kopurua dira gailendutakoak; hots, `NumTrees` eta `nMinLeafSize` elementuen arabera sailkatzaile ezberdinak eraiki dira hurrenez hurren. Horrez gain, algoritmoen garapenerako erabilitako seinale segmentuen laginen iraupena, alde batetik, 8 segundotakoak, eta bestetik, 4 segundotakoak izan dira, iraupen ezberdinen aldeak eragin ditzakeen kalte zein onurak aztertzea posible izanik. Hori horrela, lehendabizi, 8 segundotako seinale segmentuekin erdietsitako emaitzak aztertu eta, hauetan oinarriturik, eraginkorrena zein den aukeratuko da. Ondoren, 4 segundotako seinale segmentuekin lortutako algoritmoekin prozedura berdina jarraituko da, amaieran, gailendutako bi algoritmoen artean emaitza hobekien aurkezten dituen gailenduz.

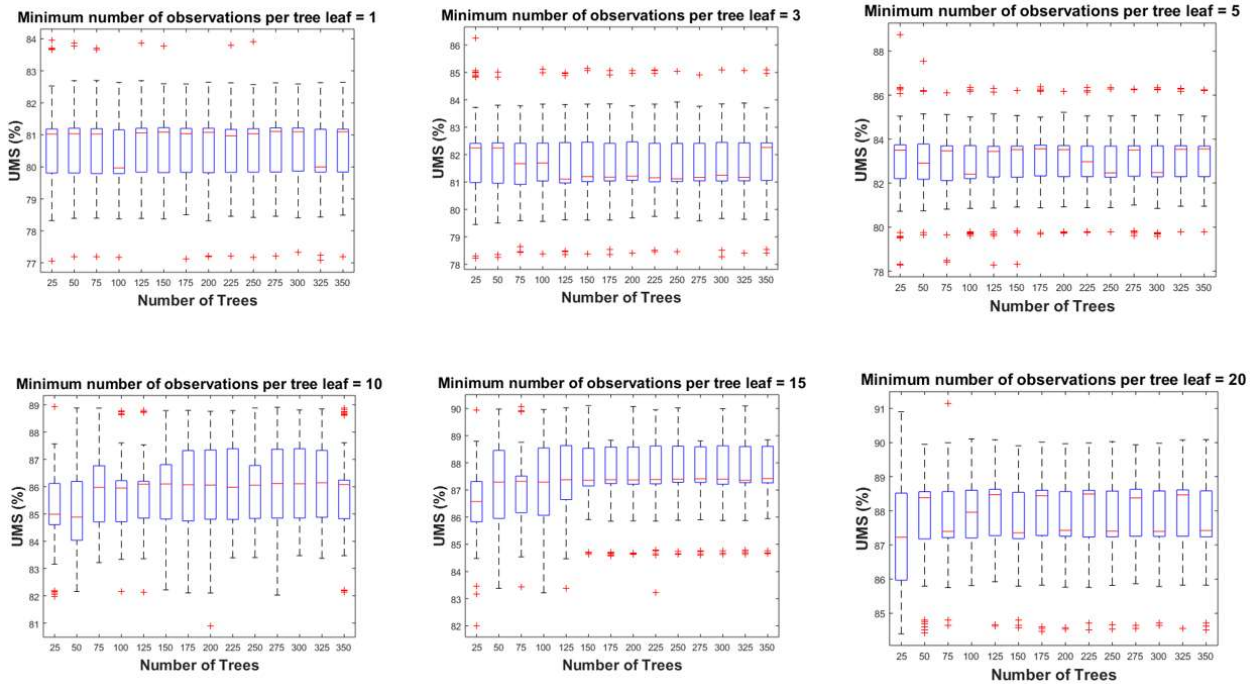
Aurkeztuko diren emaitza guztiak, Matlab-ek eskaintzen duen Box Plot funtzioan funtsaturik daude. Box Plot delakoek, kaxa diagrama gisa ere ezagunak, aztergai diren datuen sakabanaketa eta simetria bezalako ezaugarri garrantzitsu ezberdinen aldi bereko analisia ahalbidetzen dute. Izan ere, aipaturiko aztergai diren datuen sakabanaketa normala ez denean edota guztiz simetrikoa ez denean, aldagaien interpretazioa batez bestekoaren zein desbideratze estandarren arabera burutzea, hankasartzea izaten da, ez baitute elementua behar bezala deskribatzen. Kaxa diagramek, datuen egoera hobeto aurkezten duten oinarritzko hiru pertzentilak (P_{25} , P_{50} eta P_{75}), balio maximo eta minimoak eta ohikoak ez diren balioak (outlier gisa ere ezagunak) modu erraz batean bistaritzen dituzte, egoera desberdinen azterketarako eta ondorioetara ailegatzeko oso lagungarriak gertatuz.

I.I. t = 8 segundotako seinale segmentuetan oinarritutako emaitzak

Askotan azaldu denez, algoritmo eraginkorrenaren hautaketarako UMS eta MulAcc metrikak erabili dira. UMSak, jada lehenago deskribatu denez, sailkatzen ditugun erritmo klaseei pisu berdina emanez, algoritmoak aurkezten duen batez besteko sentikortasuna adierazten du; hau da, algoritmoak berpizte erritmoa ondo sailkatzeko duen batez besteko gaitasuna ematen du. MulAcc parametroak, ordea, sailkatzaileak bere osotasunean aurkezten duen doiketa, zehaztasuna neurtzen du. Horrela, bada, lehendabizi UMS emaitzak aztertu, eta jarraian, MulAcc emaitzak sakon aztertuko dira.

Aipatutakoaz gainera, zuhaitzak izan ditzakeen nodo kopuru zehatz bakoitzeko, zuhaitz kopuruaren arabera UMS eta MulAcc emaitza grafikoak, eta zuhaitz kopuru zehatz bakoitzeko, zuhaitz bakoitzak izan ditzakeen nodo kopuruaren arabera UMS eta MulAcc emaitza grafikoak aurkeztuko dira.

Jarraian, zuhaitzeko nodo kopurua zehazturik zuhaitz kopuruaren arabera UMS emaitzak ehunekotan bistaratu dira:



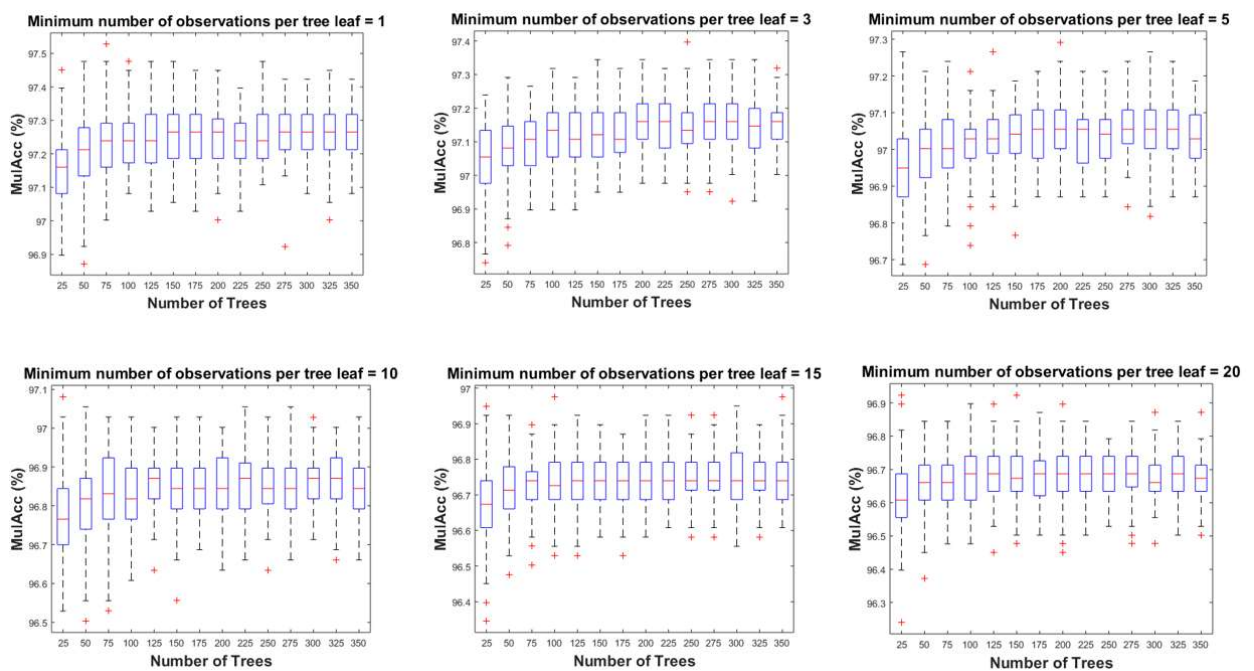
Irudia 30. Zuhaitz kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan ($t=8seg$)

Lehenengo, grafikak nola interpretatu azalduko da. Goiko grafiketan ikusten diren kaxek, aztergai diren laginen %50a barnean hartzen dutela nabarmentzekoa da; laukizuzenaren beheko mugak, %25 balio duen pertzentilari egiten dio erreferentzia, datuen %25ak, kasu honetan, muga horretan UMSak hartzen duen balioa edo txikiagoa jasotzen dutela adieraziz. Laukizuzenaren barnean dagoen marra gorriak, medianak, hots, %50a adierazten duen pertzentilari egiten dio erreferentzia, eta kaxaren goiko mugak, %75eko pertzentila adierazten du. Laukizuzenetik kanpo dauden goiko ete beheko mugek, beraz, minimoari eta maximoari egiten diote erreferentzia.

Irudia 30ko lehenengo hiru grafikei so eginez, zuhaitz kopuruaren araberako laginen egoera eta bilakaera nahiko antzekoa eta homogeneoa dela ikus dezakegu. Hala ere, zuhaitz bakoitzeko adar bakoitzeko behaketa kopurua handitu ahala, algoritmoaren berpizte erritmoak ondo sailkatzeko gaitasuna %81.0 ingurutik %83.0 ingurura gora doala nabarmentzekoa da. Hurrengo hiru grafikoetan, zuhaitz kopuruaren araberako UMS balioek aldaketa handiagoak aurkezten dituztela ikusten da. Zuhaitz adar bakoitzeko gutxienez 10 behaketa gauzatzen dituen emaitzetan, laginen sakabanatzea handiegia dela bistakoa da, UMSak hartzen duen balio minimotik balio maximora dagoen saltoa, orokorrean, sekulakoa baita. Hurrengo grafikoko emaitzek aurkezten duten bilakaera, ordea, argiagoa da: hasieran, UMS proportzioek pixkanaka-pixkanaka gora egiten dute, sailkatzaileak 150 zuhaitz erabiltzetik aurrera emaitzak egonkortzen direlarik. Azken hauek, ordea, ezohikoak diren hainbat balio aurkezten dituzte, algoritmoetan, normalean, onuragarriak izaten ez direlarik. Azkeneko grafikoan behatuz, 25 zuhaitzetan oinarritutako sailkatzailea alde batera utziz, gainerakoek, UMS balio tarte oso antzekoak erakusten dituzte, sakabanaketan eta izaeran zenbait alde agertzen dituzten arren; 50, 125, 175, 225, 275 eta 325 zuhaitz darabilten sailkatzaileetan, laginen kontzentrazio handiena %88.3 - %88.6 tarteetako balioa aurkezten duten UMSan aurkitzen dira, erabili diren laginen %25a bertan

egonik. Laginen %50ak %86.0-%88.3 bitarteko UMS balioak hartzen dituzte, sakabanaketa nahiko hedatua aurkeztuz, eta gainerako %50ak %88.3-%90.0 bitarteko UMS balioak hartzen dituzte. 75, 150, 200, 250, 300 eta 350 zuhaitzetan oinarrituriko sailkatzaileek, kontrara, laginen %25eko kontzentrazioerik handiena, %87.1 - %87.4 bitarteko UMS balioak hartzen dituzte. Medianak %87.3 balioa harturik, UMS balio horretatik %89ko UMS baliorarte aztergai diren datuen erdia aurkitzen dira, hemen sakabanatzea gainerako datuen sakabanatzea baino askoz ere hedatuagoa izanik (laginen beste %50a %85.7-%87.3 UMS balio tartea hartzen dute). Kasu honetan, sailkatzailearen errendimenduari dagokionez, onuragarriagoak lehenengo kasuko algoritmoak lirarteke, datuen %50ak aipatutako beste algoritmoek igartzen dituzten datuen %50ak baino UMS balio altuagoak eskaintzen baitituzte. Hauek, aldiz, ohikoak ez diren balio kopuru nahiko handia aurkezten dituzte. 100 zuhaitzekin eraikitako sailkatzaileak, ahatik, ezohiko baliorik ez duen banaketa normal eta simetrikoa erakusten du.

MulAcc parametroari dagokionez, Irudia 31an zuhaitz kopuruaren arabera eraikitako sailkatzaile ezberdinen emaitzak eskuragarri daude. Jakinaenez, MulAcc parametroak sailkatzaileak bere osotasunean duen zehaztasuna aurkezten du; metrika honek, ordea, UMS metrikak ez bezala, erritmoen proportzioa kontuan hartzen ez duela, hau da, klaseen desoreka alde batera uzten duela azpimarratzekoa da. Gure kasuan, algoritmoaren eraikuntzarako erabiliko datu-baseak biltzen dituen erritmoen proportzioetan orekarik aurkezten ez duenez gero, MulAcc parametro honek ematen dituen emaitzak kontuan hartuko badira ere, bereziki UMS parametroaren emaitzetan funtsatuko gara.

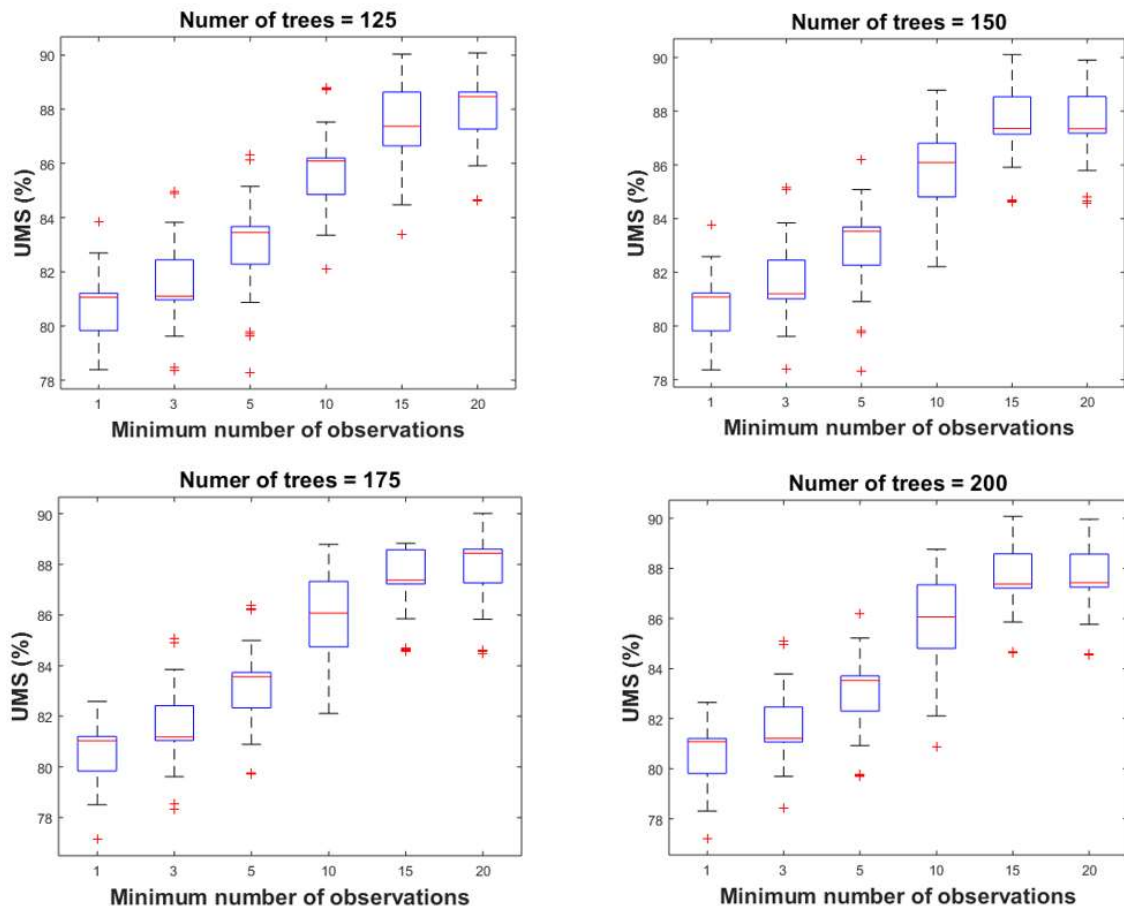


Irudia 31. Zuhaitz kopuruaren araberako algoritmoek aurkezten dituzten MulAcc emaitzak ehunekotan

Lortutako MulAcc emaitzak behatuz, UMS emaitzekin ez bezala, zuhaitzeko adar bakoitzeko behaketa kopurua handitu ahala, MulAcc balioak behera egiten du, aldea oso handia ez den arren (%97.2ko mediana batetik %96.6ko mediana batera egiten du jauzi gutxi gorabehera). Balio egonkorragoak, UMS emaitzekin bezala, algoritmo-sailkatzaileak 100-125 zuhaitz erabiltzetik aurrera agertzen direla ikus dezakegu, emaitzak nahiko antzerakoak direlarik. Sailkatzaileerik eraginkorrena eta errendimendurik handiena aurkezten duena aukeratzeko irizpideen artean, sentikortasun altua eta

zehaztasuna izateaz gain, ahalik eta kostu konputazional txikienarekin lortzea ere kontuan hartzekoa da. Horrez gain, ohiz kanpoko laginen balioak eta hauen kopurua ere aintzat hartzeko parametroa dugu; askotan, izan ere, aipaturiko ohikoak ez diren balioak barnean dituzten datuen multzoetatik eratorritako estatistikak ez dira fidagarriak izaten orokorrean. Azkeneko honi loturik, behaketa kopurua handitu ahala ohikoak ez diren balioen kopurua ere handituz doala azpimarratzekoa da.

Beste ikuspuntu bat izatearren, UMS eta MulAcc balioek aurkezten duten izaera, sailkatzaileek burutzen dituzten gutxienezko behaketa kopuruaren arabera aurkeztuko dira. Orain arte eskuraturiko emaitzak oinarri harturik, badirudi 200 zuhaitzetatik gora eraikitako sailkatzaileen emaitzek oso izaera antzekoa dutela; horrez gain, 125tik beherako zuhaitz kopuruarekin eraikitako sailkatzaileetan, metriken hobekuntza ikusi da. Hori horrela, 125, 150, 175 eta 200 zuhaitzetan funtsaturiko algoritmoetan oinarrituko gara lehenengo, ondoren, 225, 250, 275 eta 300 zuhaitzetan oinarrituriko algoritmoen emaitzek dituzten aldaketa txiek ere eraginik duten aztertuz.



Irudia 32. Behaketa kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan 125, 150, 175 eta 200 zuhaitz erabilita hurrenez hurren ($t=8\text{seg}$)

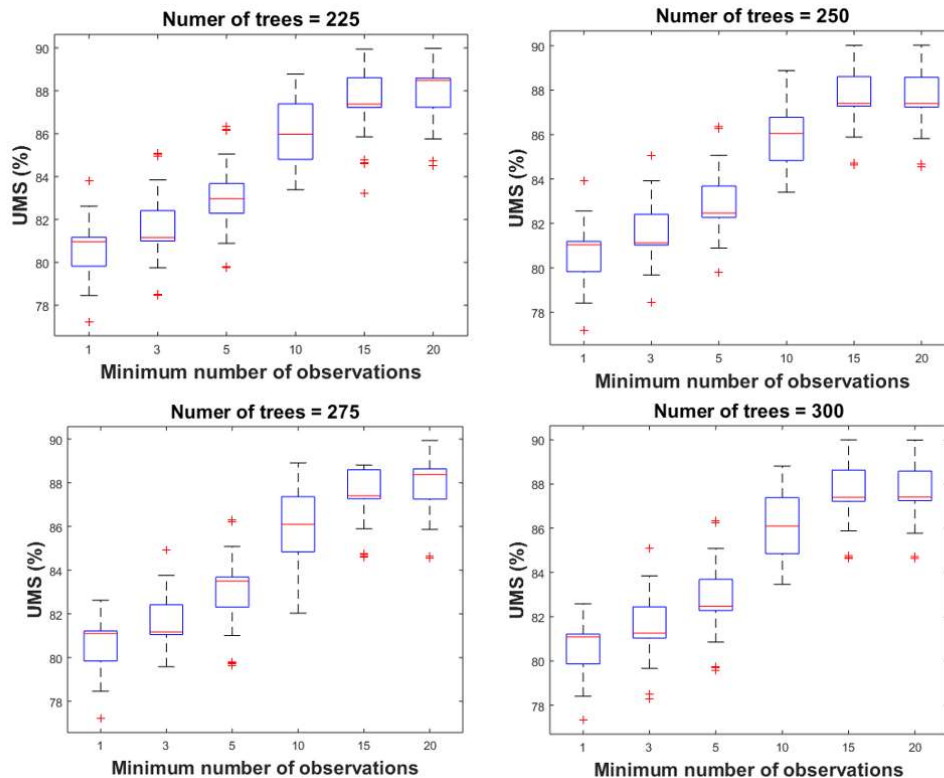
Irudia 32ri so eginez, sailkatzaile bakoitzean erabilitako behaketa kopurua igo ahala UMS balioaren bilakaera gorakorra dela agerikoa da, 15 behaketetatik aurrera oinarrituriko UMS balioak egonkortzen direlarik. Emaitzak agertzen dituzten lau grafikoetatik, eraginkorrenak 150, 175 eta 200 zuhaitzetan

oinarritutako sailkatzaileetan aurkitzen dira. 125 zuhaitzetan oinarritutako sailkatzailea, aurretik ikusi den MulAcc balioek aurkezten zuten zehaztasun maila txikiago eta sakabanatuago bat izateagatik, baztertu da. Azpimarraturiko azkeneko hiru kasu hauetan, gutxienezko 10, 15 eta 20 behaketa burutzen dituzten sailkatzaileen UMS balio tartetan alde esanguratsurik aurkezten ez dutela bistakoa da.

Gutxienezko 10 behaketa gauzaten dituen algoritmoaren kasuan, alde gutxirekin, baina emaitza hobek, 175 edota 200 zuhaitzetan oinarritutako sailkatzaileek eskaintzen dituzte. Biek aurkezten dituzten emaitzen banaketa berdina dela esan genezake: aztergai diren laginen %50, %82.1-%86.1 tarteko UMS balioen artean aurkitzen dira, hauen artean %25, %84.7-%86.1 tarteko UMS balioen artean kontzentratuak dagoela. Gainerako laginek, hau da, guztiz simetrikoak diren datuen beste %50ak, %86.1-%88.7 tarteko UMS balioak hartzen dituzte. 200 zuhaitzekin eraikitako algoritmoak, ordea, besteak ez duen UMS balio baxuago bat hartzen duen outlier bat agertzen du, bien artean, beraz, errentagarriena eta efizienteena 175 zuhaitzetan oinarrituriko algoritmoa izango litzatekeelarik.

Gutxienezko 15 eta 20 behaketetako emaitzei dagokienez, bien arteko aldeak oso txikiak dira; hala ere, emaitzarik eraginkorrenak 175 zuhaitz eta zuhaitz bakoitzeko gutxienezko behaketa kopurua 20 izanik sorturiko sailkatzailearenak dira, laginen %50 balio altuenetan kontzentratuak baitaude. Outlier balio gutxi batzuk aurkezten baditu ere, hauek beste sailkatzaile batzuek eskaintzen dituzten balioak hartzen dituzte, %84.5 inguruko UMS balio esanguratsua izanik, eta hortaz, guztiz arbuigarriak izanik.

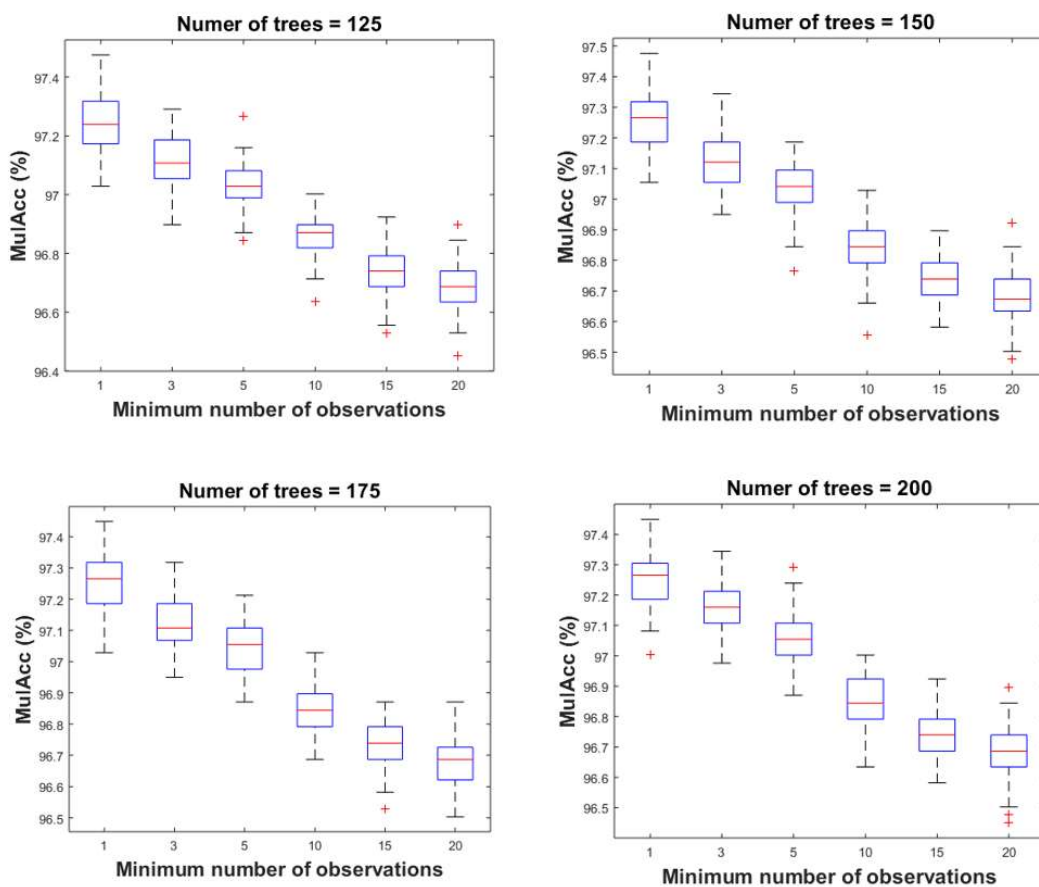
Orain, begi bistaz egoera egonkorragoa aurkezten duten 225, 250, 275 eta 300 zuhaitzekin eraikitako sailkatzaileen UMS emaitzak aztertuko dira, horrela, alde nabarmenik dagoen ala ez baieztatzeko:



Irudia 33. Behaketa kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan 225, 250, 275 eta 300 zuhaitz erabilita hurrenez hurren ($t=8\text{seg}$)

Aurreko egoerekin ikus dezakegun alde bakarra, gutxienez 10 behaketa burutzen dituen algoritmoei dagokienez, 225, 250 eta 300 zuhaitzetan oinarritutakoek, UMS minimo zertxobait altuago bat definitzen dutela da: %82.1etik %83.4rako igoera gutxi gorabehera; dena den, gutxienezko igoera hau ematen duen algoritmoak beharrezkoak dituen zuhaitz kopuru handiago horietara igotzeak zentzu handirik ez du, prozesaketa denboraren igoera merezi ez baitu. Aztertutako beste bi egoerei dagokienez (15 eta 20 behaketetan oinarritutakoak), emaitzak berdinak izaten jarraitzen dute, eta hortaz, 175 zuhaitzekin eraikitako sailkatzailea eraginkorragoa izaten jarraitzen du, honek behar duen prozesaketa denbora txikiagoa baita.

Jarraian, gutxienezko behaketen araberako MulAcc emaitzak bistaratu eta aztertuko dira:



Irudia 34. Behaketa kopuruaren araberako algoritmoek aurkezten dituzten MulAcc emaitzak ehunekotan 125, 150, 175 eta 200 zuhaitz erabilita hurrenez hurren ($t=8\text{seg}$)

MulAcc parametroak, badakigunez, datu-basea osatzen duen lagin bakoitza pisu berdinarekin ebaluatu egiten du. Geuk erabilitako datu-baseak aurkezten duen erritmoen proportzio desoreka dela medio, parametro honek eskaintzen digun sistemaren guztizko doitasunari buruzko informazioa, ez da horren zehatza ezta fidagarria ere; izan ere, laginen erdia baino gehiago, esaterako, AS erritmoari egokitzen zaio, eta beraz, honekin bat egiteko probabilitatea oso handia da. Gainera, Irudia 34ak bistaratzen dituen grafikoei so eginez, behaketa kopurua zenbat eta handiagoa izan MulAcc

zehaztasun balioak behera egiten duela agerikoa da; alabaina, eta lehenago ere aipatu denez, doitasun balioen aldeak oso txikiak dira.

Alderaturiko gutxienezko 20 behaketekin eraikitako eta 175 eta 200 zuhaitzetan oinarrituriko sailkatzaileen zehaztasun mailari dagokionez, biek aurkezten duten izaera nahiko antzekoa da; hala ere, zenbait alde ere agertzen dituzte: lehenengo sailkatzaileak (175 zuhaitzetan oinarritutako) ez du ez ohiko baliorik aurkezten, bigarrenak (200 zuhaitzetan oinarritutako) 3 outlier aurkezten dituelarik. Horrez gain, lehenengo sailkatzaileak zehaztasun balio zertxobait altuagoak ere agertzen ditu.

Orain arte aipatutako guztia aintzakotzat harturik, $t=8$ segundotako seinale segmentuak erabilia eta bereziki UMS balioetan oinarriturik, algoritmo-sailkatzailerik eraginkorrena 175 zuhaitz eta gutxienezko 20 behaketetan oinarrituriko sailkatzailea dugula ondoriozta dezakegu. Egia da, baldintza hauetan funtsaturiko algoritmoaren zehaztasun maila, baxuenetarikoa dela: datuen %50a, alde batetik, %96.5 eta %96.7 tarteko zehaztasuna aurkeztu, eta bestetik, gainerako datuek %96.7 eta %96.9 tarteko zehaztasunean aurkitzen baitira. Kasu honetan, ordea, klaseen desorekarekiko sentikorra ez den UMS metrika da batez ere aintzat hartzea erabaki dena; izan ere, MulAcc balioak, ondo sailkatzen dituen erritmoen inguruko informaziorik eman ez, eta algoritmoarekiko errendimendu eta eraginkortasunarekiko eskaintzen duen fidagarritasuna txikiagoa baita.

Hautaturiko algoritmoaren ebaluazio metriken laburpena biltzen duen taula jarraian eskuragarri dago:

Taula 35. 8 segundotako seinale segmentuetan oinarrituriko 175 zuhaitz eta gutxienezko behaketa kopurua 20 dituen algoritmoaren UMS eta MulAcc metriken balioen laburpen taula

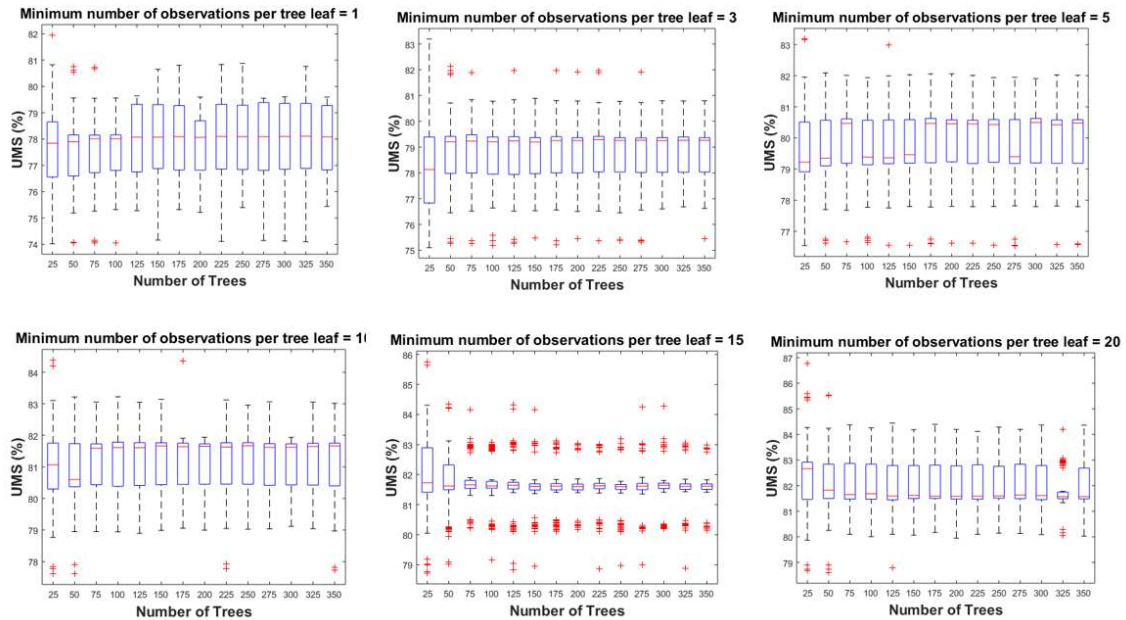
METRIKA	UMS (%)	MulAcc (%)
MIN	85.83	96.50
MAX	90.01	96.87
P ₂₅	87.28	96.62
P ₅₀	88.44	96.69
P ₇₅	88.60	96.73
OUTLIER_1	84.61	-
OUTLIER_2	84.57	-
	84.47	-

I.II. $t = 4$ segundotako seinale segmentuetan oinarritutako emaitzak

Zortzi segundotako segmentuen bitartez lorturiko algoritmoak azterturik, 4 segundotako segmentuetan oinarritutakoak dira jarraian deskribatu eta analizatuko direnak. Lehendabizi, aurreko kasuan moduan, algoritmoek darabilten behaketen araberako UMS sentikortasun maila neurtu eta azaltzen duten emaitzen grafikoak analizatuko dira.

Irudia 35ari so eginez, 4 segundotako tartekak erabilia lortutako algoritmoen UMS sentikortasun mailak 8 segundotako tartekak erabilia eraikitako algoritmoek aurkezten duten UMS sentikortasun mailak baino baxuagoak direla ikus daiteke. Emaitzak, begi bistaz, aurreko kasuko emaitzak baino zertxobait kaskarragoak diren arren, 4 segundotako segmentuetan oinarrituriko algoritmoen artean eraginkorrena ere aukeratuko da. Algoritmoaren hautaketarako, lehenago esan den bezala, UMS parametroa izango da batez ere aintzat hartuko dena, honek laginen erritmoen desoreka proportzioak kontuan hartzen dituela badakigu eta. Ildo honetatik, gutxienez 10 behaketa edota 20 behaketa

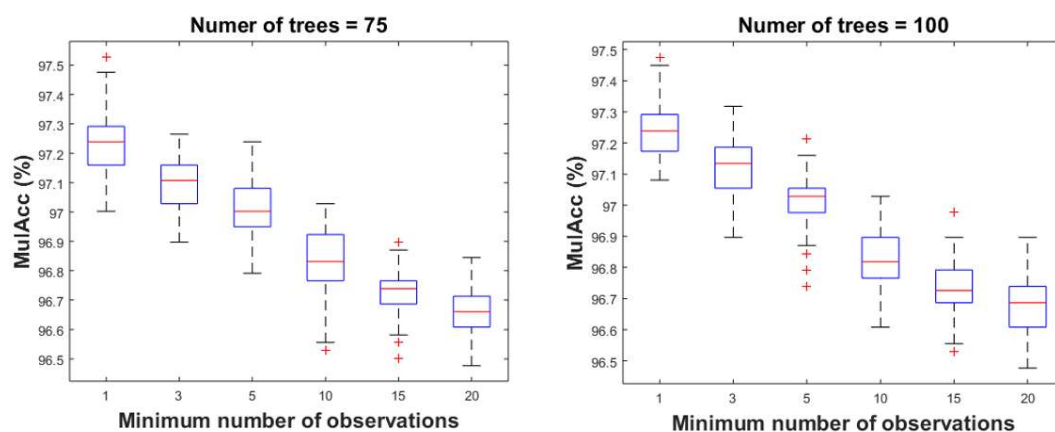
burutzen dituzten sailkatzaileak dira gailenduko ditugunak, 15 behaketetan oinarrituriko algoritmoek egonkortasunik aurkezten ez baitute.



Irudia 35. Zuhaitz kopuruaren araberako algoritmoek aurkezten dituzten UMS emaitzak ehunekotan ($t=4\text{seg}$)

Gutxienez 10 behaketa burutzen dituen algoritmoen artean, aurkezten duten UMS balioen sakabanaketa nahiko antzerakoa dela ikusita, prozesaketa denbora gutxiago behar duen 100 zuhaitzetan oinarritutako sailkatzailea erangikorrena eta errendimendurik altuena agertzen duela esango genuke: laginen %50a %81.6 eta %83.2 UMS balioen artean aurkitzen dira, gainerako laginen %50a %78.9 eta %81.6 UMS balioen artean aurkitzen delarik. Gutxienez 20 behaketa burutzen dituen algoritmoen artean ere, UMS balioen sakabanaketa oso antzekoa da. Kasu honetan, beraz, 75 zuhaitzetan oinarritutako sailkatzailea erangikorrena eta errendimendurik altuena agertzen duela esango genuke, honek ere, beste sailkatzaileekin alderatuz, behar duen prozesaketa denbora txikiagoa delako. Kasu honi dagokionez, laginen %50a %81.7 eta %84.4 UMS balioen artean aurkitzen da, gainerako laginen %50a %80.1 eta %81.7 UMS balioen artean aurkitzen delarik.

Aipaturiko bi kasu hauen artean zein aukeratu erabakitzeko, algoritmoak, bakoitzak bere aldetik, osotasunean aurkezten duen MulAcc zehaztasun maila aztertuko da jarraian:



Irudia 36. Behaketa kopuruaren araberako 75 eta 100 zuhaitzetan oinarrituriko algoritmoek aurkezten dituzten MulAcc emaitzak ehunekotan ($t=4\text{seg}$)

Irudia 36a behatuz, bai 75 zuhaitzekin eta gutxienezko 20 behaketekin garaturiko algoritmoak, baita 100 zuhaitzekin eta gutxienezko 10 behaketekin eraikitako algoritmoak ere, MulAcc sakabanaketa nahiko simetrikoa erakusten dute.

Lehenengo kasuari dagokionez (75 zuhaitz eta 20 behaketa), honek aurkezten dituen MulAcc balioak hurrengoak dira: datuen %50ak %96.5 eta %96.7 mailako zehaztasuna, eta gainerako datuen %50ak %96.7 eta %96.8 bitarteko MulAcc balioak aurkezten ditu. Gainera, balioen %25a %96.6 eta %96.7 bitarteko zehaztasun balioetan kontzentratuak aurkitzen da, laginen beste %25a %96.7 eta %96.7 bitarteko zehaztasun balioetan kontzentratuak aurkitzen delarik.

Bigarren kasuari dagokionez (100 zuhaitz eta 10 behaketa), ordea, honek aurkezten dituen MulAcc balioak hurrengoak dira: datuen %50ak %96.6 eta %96.8 mailako zehaztasuna, eta gainerako datuen %50ak %96.8 eta %97.0 bitarteko MulAcc balioak aurkezten ditu. Gainera, balioen %25a %96.8 eta %96.8 bitarteko zehaztasun balioetan kontzentratuak aurkitzen da, laginen beste %25a %96.8 eta %96.9 bitarteko zehaztasun balioetan kontzentratuak aurkitzen delarik.

Grafikoko emaitzak zein aipaturiko emaitzak behatuz, argi dago, doitasun eta zehaztasun hobekak bigarren kasuak eskaintzen dituela. UMS zein MulAcc balioen arteko aldea, bi kasuetan, oso txikia da; hala ere, lehenago aipatu izan denez, datu-baseak biltzen dituen erritmoen desoreka proportzioa dela medio, UMS sentikortasun parametroari, MulAcc parametroak eskaintzen duen doitasun balioari baino garrantzi handiagoa emango zaio proiektuan. Hori horrela, hautatutako 4 segundotako seinale segmentuetan oinarritutako algoritmoa 75 zuhaitz eta gutxienezko 20 behaketekin eraikitakoa da. Jarraian, azkeneko algoritmo honi dagokion parametroen laburpen taula aurkezten da:

Taula 36. 4 segundotako seinale segmentuetan oinarrituriko 75 zuhaitz eta gutxienezko behaketa kopurua 20 dituen algoritmoaren UMS eta MulAcc metriken balioen laburpen taula ($t=4\text{seg}$)

METRIKA	UMS (%)	MulAcc (%)
MIN	80.10	96.45
MAX	84.37	96.84
P ₂₅	81.48	96.61
P ₅₀	81.65	96.66
P ₇₅	82.87	96.71

Emaitzak aztertuz, 8 segundotako segmentuetan funtsaturiko algoritmoaren emaitzei dagokien errendimendu eta eraginkortasunari dagokionez, agerikoa da emaitzak askoz hobekak direla. Hori,

algoritmoaren garapenerako burutu beharreko berebiziko garrantzia duen ezaugarrien erauzketan duen eraginagatik da batez ere; berpizte erritmo bat deskribatzen duen seinale segmentua iraupen luzeagoa izateagatik, seinalea hobeto deskribatzen duten berariazko ezaugarriak lortzea eragiten baitu.

Honekin guztiarekin, Random Forest algoritmoa aplikaturik ikasketa automatikoan oinarritutako berpizte bihotz erritmoen sailkatzailek eraginkorrena, batez beste %96.5ko zehaztasuna eta %88.0ko sentikortasuna eskaintzen duen 175 zuhaitzekin eta gutxienezko 20 behaketekin gauzaturik $t=8$ segundotako seinale segmentuetan oinarrituriko sailkatzailea dela ondoriozta dezakegu.