



**UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI"
DIN IAȘI**



Funcții și arhitecturi 6G aplicate în rețelele viitoare de telecomunicații

Marius Iordache

Conducător de doctorat: Prof. dr. ing. Ion Bogdan

IAȘI, 2025

UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI" DIN IAȘI
R E C T O R A T U L

Către

Vă facem cunoscut că, în ziua de _____ la ora _____ în
_____, va avea loc
susținerea publică a tezei de doctorat intitulată:

"Funcții și arhitecturi 6G aplicate în rețelele viitoare de telecomunicații"

elaborată de doamna/ domnul **drd. ing. IORDACHE M. MARIUS** în vederea conferirii titlului
științific de doctor.

Comisia de doctorat este alcătuită din:

- | | |
|---|------------------------|
| 1. (funcție didactică, nume, prenume, universitatea de proveniență) | președinte |
| 2. (funcție didactică, nume, prenume, universitatea de proveniență) | conducător de doctorat |
| 3. (funcție didactică, nume, prenume, universitatea de proveniență) | referent oficial |
| 4. (funcție didactică, nume, prenume, universitatea de proveniență) | referent oficial |
| 5. (funcție didactică, nume, prenume, universitatea de proveniență) | referent oficial |

Cu această ocazie vă invităm să participați la susținerea publică a tezei de
doctorat.

RECTOR,

.....

Secretar universitate,

.....

*pentru susținerile online se va preciza link-ul și soluția de software

În primul rând, doresc să adresez cele mai sincere mulțumiri domnului Prof. univ. dr. ing. Ion Bogdan, coordonatorul lucrării de doctorat, pentru îndrumarea atentă, profesionalismul remarcabil și sprijinul constant oferit pe parcursul întregii cercetări. Răbdarea, exigența științifică și încurajările domniei sale au avut un rol esențial în dezvoltarea acestei teze. Sub îndrumarea sa, am avut ocazia să îmi dezvolt gândirea critică, să abordez tematici complexe și să ating un nivel înalt de rigoare academică.

Totodată, îmi exprim profunda recunoștință față de referenții științifici ai tezei, pentru timpul acordat evaluării acestei lucrări, pentru observațiile pertinente și pentru contribuția valoroasă la îmbunătățirea conținutului științific.

Mulțumiri deosebite adresez și cadrelor didactice ale Universității Tehnice „Gheorghe Asachi” din Iași, care, prin profesionalismul și dedicarea lor, au contribuit la dezvoltarea mea intelectuală și științifică.

Totodată, adresez mulțumiri și companiei Orange România, pentru oportunitatea oferită de a desfășura partea practică a cercetării în cadrul unei organizații de referință în domeniul telecomunicațiilor. Accesul la infrastructura reală, deschiderea specialiștilor din companie și sprijinul acordat în procesul de colectare și analiză a datelor au contribuit în mod semnificativ la aplicabilitatea și relevanța practică a acestei teze.

Această lucrare este rezultatul unui efort colectiv, iar tuturor celor care m-au sprijinit în acest parcurs le sunt profund recunoscător.

Cuprins

1. Introducere	7
1.1 Scopul cercetării	7
1.2 Structura și conținutul tezei de doctorat.....	8
2. Evoluția sistemelor de telecomunicații	10
2.1 Arhitectura sistemului 5G	10
2.2 Evoluția arhitecturii de rețelei	11
2.3 Evoluția sistemului telecom	12
3. Funcționalități de administrare automată a rețelei.....	13
3.1 Abordări standardizate în industrie.....	13
3.2 Proiecte de cercetare europene	15
3.3 Managementul slice-urilor de rețea	15
3.4 Arhitectura propusă pentru framework-ul AI/ML.....	16
4. Tehnici si Algoritmi	19
4.1 AI/ML intro in telecom systems	19
4.2 Premise necesare pentru realizarea AI/ML	19
4.3 Soluție de achiziție a datelor	22
4.4 Sistem de management al rețelei bazat pe inteligență artificială.....	24
4.5 Implementarea AI/ML și rezultatele inițiale	26
4.6 Seturi de date utilizate.....	27
4.7 Dezvoltarea algoritmului.....	30
5. Rezultate si Analiză	33
5.1 Mediul de testare	33
5.2 Scenariu de configurare	34
5.3 Implementare unui model AI/ML	34

5.4	Concluzii.....	36
6.	Concluzii finale, contribuții proprii și direcții de cercetare viitoare	38
6.1	Contribuție proprie.....	38
6.2	Lista Publicatiilor.....	40
6.3	Direcții viitoare.....	41
	References	42

1. Introducere

Industria telecomunicațiilor se află într-o fază de transformare, încercând să migreze de la implementările și perspectivele tradiționale către o viziune nouă bazată pe AI/ML. Soluțiile sunt fundamentate pe câteva inițiative cheie, precum rețelele native în cloud, AI/ML, operare complet automatizate pentru implementări și configurări, rețelele bazate pe intenție și procesele de auto-configurare. Specificitatea industriei telecomunicațiilor constă în faptul că, fiind considerate infrastructuri critice datorită serviciilor de comunicații furnizate (inclusiv cele de urgență), reglementările sunt foarte stricte. Datorită acestor aspecte, evoluția va avea loc în etape, pe parcursul mai multor ani. Lucrarea pe care am realizat-o include analize și cercetări acoperind aspectele tehnologice vizate, implementările existente în raport cu rețelele 5G actuale, transformarea pe termen scurt a rețelelor și evoluția pe termen lung.

1.1 Scopul cercetării

Industria telecomunicațiilor este un sector tehnologic care evoluează, încercând să migreze de la implementările și perspectivele tradiționale către viziunea nouă bazată pe inteligență artificială (AI) și învățare automată (ML). Soluțiile se bazează pe cadre complet automatizate pentru operare, implementări și configurări, rețele ghidate de intenție (intent) și procese de autoreparare. Datorită specificului industriei telecom, evoluția va avea loc în etape, pe parcursul mai multor ani.

Activitatea pe care am desfășurat-o include analize și cercetări care acoperă aspectele tehnologice vizate, implementările existente în raport cu rețelele 5G actuale, transformarea rețelelor pe termen scurt și evoluția pe termen lung. Aceste aspecte se suprapun în principal în două direcții:

- încotro se îndreaptă rețelele pentru a valorifica tehnologiile existente sau potențialele capacități viitoare, și
- care sunt cerințele de afaceri, noile cazuri de utilizare și nevoile de comunicare care declanșează această evoluție.

Aceste elemente reprezintă baza fundamentală a evoluției viitoare a telecomunicațiilor, întrucât, în cadrul acestor concepte inovatoare, nu putem neglija principalul factor al oricărei activități comerciale: sustenabilitatea afacerii și viabilitatea financiară.

Mai multe tendințe tehnologice au fost identificate pentru dezvoltare în cadrul unor faze diferite de implementare; acestea pot fi exprimate ca arii de tendință și factori tehnologici facilitatori pentru stabilirea rețelelor viitorului. Tendințele sunt aliniate cu scopul lucrării de față.

Industria telecomunicațiilor evoluează către o rețea inteligentă, modulară, programabilă și sustenabilă, sprijinită de AI. Este evident că rolul inteligenței artificiale (AI) în rețelele de telecomunicații ale viitorului este esențial. AI/ML (învățarea automată) va fi o componentă software integrată nativ, având ca scop automatizarea, optimizarea și autogestionarea tuturor nivelelor rețelei. Inteligența artificială va susține comunicațiile bazate pe intenții, prin traducerea cerințelor utilizatorilor în configurații de rețea, folosind modele lingvistice de mari dimensiuni (LLMs) pentru interpretarea situațiilor complexe. AI va controla resursele rețelei pentru a asigura optimizarea performanței, eficiența costurilor și interacțiunea cu soluțiile de orchestrare, gestionarea traficului de transport și alocarea spectrului RAN. Un exemplu aplicabil este activarea dinamică și configurarea automată a "slice-urilor" de rețea.

Inteligența artificială îmbunătățește timpul de livrare a serviciilor și permite operatorilor de rețea să ofere partenerilor externi servicii îmbunătățite, prin mentenanță predictivă și gestionarea automată a defectelor. Totodată, AI este factorul facilitator pentru un model digital al rețelei de comunicații (digital network twins), contribuind la testarea, optimizarea și analiza predictivă în managementul rețelei, prin simularea funcțiilor de rețea (Reprezentare digitală a rețelei) înainte de implementarea acestora în scenarii reale.

1.2 Structura și conținutul tezei de doctorat

Primul capitol prezintă o sinteză a tezei, evidențiind rolul AI/ML în evoluția rețelelor de telecomunicații. Totodată, el descrie tranziția de la 5G la 6G, subliniind necesitatea remodelării arhitecturii rețelelor pentru a integra AI/ML nativ. Se introduce conceptul de management de tip zero-touch și automatizare, prezentând provocările și soluțiile propuse pentru implementarea AI/ML în rețelele viitoare. În capitolul al doilea se prezintă contextul general al cercetării, prezentând evoluția istorică a telecomunicațiilor și accelerarea schimbărilor tehnologice. El identifică tendințele actuale în industria telecom, atât din perspectivă de business (cazuri de utilizare 6G), cât și tehnologică (AI nativ, edge computing, etc.). Se definește problema principală - adoptarea lentă a AI/ML în managementul rețelelor - și prezintă metodologia de cercetare, obiectivele și contribuțiile așteptate.

Capitolul trei analizează evoluția sistemelor de telecomunicații, de la arhitectura actuală 5G bazată pe servicii (SBA) către viziunea 6G. Sunt descrise componentele principale ale rețelelor 5G, virtualizarea funcțiilor de rețea și conceptul de “network slicing”. Prezintă etapele de evoluție ale rețelelor telecom (Telco 1.0-4.0), de la infrastructuri fizice la rețele autonome bazate pe AI, subliniind transformările arhitecturale și operaționale necesare pentru viitoarele rețele 6G.

În următoarele patru capitole sunt prezentate rezultatele obținute cu metodele dezvoltate și implementate.

Capitolul 4, investighează caracteristici de gestionare automată a rețelei, explorează conceptul de management zero-touch pentru rețelele 6G, analizând abordările standardizate (3GPP, ETSI ZSM, TM Forum) și proiectele de cercetare europene (ADROIT-6G, 6G-INTENSE). Prezintă conceptele de management bazat pe intenții (intent-based), bucle de control închise (closed-loop) și automatizare. Propune o arhitectură inovatoare pentru 6G care integrează nativ AI/ML la toate nivelurile, pentru a permite managementul autonom al rețelei și optimizarea resurselor.

În capitolul cinci, se prezintă algoritmi evaluați (GRU, LSTM, Random Forests, Gradient Boosting) și se justifică selecția GRU pentru predicția performanței rețelei. Descrie sistemul de achiziție a datelor implementat pentru colectarea metricilor din rețea și procesul de dezvoltare a modelului AI/ML, incluzând preprocesarea datelor, antrenarea și evaluarea. Prezintă rezultatele preliminare ale modelului pe date reale din rețea, introduce conceptele fundamentale de AI/ML și diferențele față de configurarea clasică.

Capitolul 6, Rezultate si Analiză, se descrie mediul de testare utilizat (Orange Romania 5GLAB) și configurația scenariului de test, incluzând rețeaua 5G SA, network slicing și aplicațiile de monitorizare. Prezintă implementarea practică a modelului AI/ML pentru predicția degradării serviciilor și reconfigurarea automată a slice-urilor de rețea. Demonstrează funcționalitatea sistemului propus într-un scenariu real, unde AI/ML detectează potențiale probleme de performanță și declanșează automat ajustări de QoS pentru a menține SLA-urile.

În final, capitolul șapte sintetizează contribuțiile principale ale cercetării: arhitectura 6G cu AI/ML nativ, framework-ul pentru managementul zero-touch și demonstrația practică a automatizării bazate pe AI/ML.

2. Evoluția sistemelor de telecomunicații

Capitolul analizează evoluția sistemelor de telecomunicații, de la arhitectura actuală 5G către viziunea 6G[1], evidențiind transformările arhitecturale necesare pentru a susține viitoarele aplicații și servicii.

2.1 Arhitectura sistemului 5G

Secțiunea prezintă arhitectura de bază a sistemului 5G, care a evoluat către o Arhitectură Bazată pe Servicii (SBA). Sistemul 5G este proiectat pentru conectivitate de date și servicii, permițând implementări prin Virtualizarea Funcțiilor de Rețea și Rețele Definite prin Software. Arhitectura 5G include rețele virtualizate, separarea planului de utilizator (UP) de planul de control (CP), și interfețe bazate pe servicii pentru comunicarea între funcțiile de rețea.

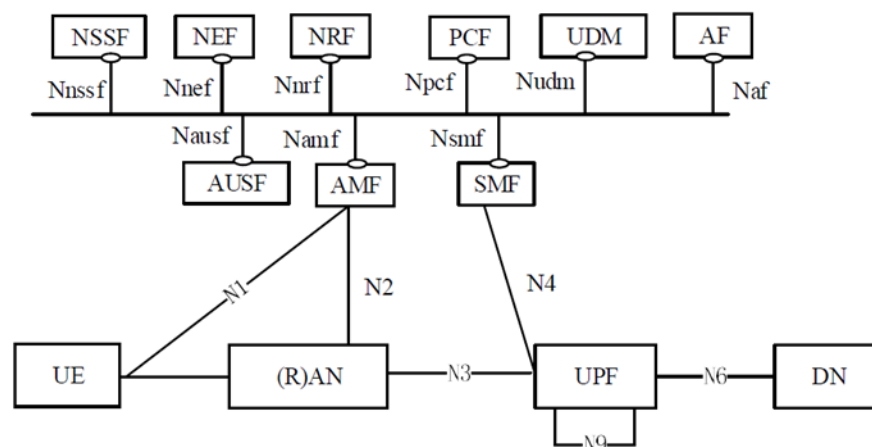


Figura 2-1 Arhitectura sistemelor 5G

5G include componente funcționale, cum ar fi AMF (funcția de gestionare a accesului și mobilității), UPF (funcții ale planului utilizatorului) și SMF (funcția de gestionare a sesiunii), alături de alte funcții, așa cum este ilustrat în Figura 2-1.

Arhitectura nucleului 5G (5G Core) reprezintă o evoluție a sistemelor, introducând schimbări majore prin adoptarea arhitecturii bazate pe servicii (Service-Based Architecture – SBA). În terminologia SBA, componentele arhitecturale principale sunt serviciile, care sunt utilizate pentru a executa diferite seturi de funcții, având permisiunea de a accesa reciproc serviciile celorlalte componente. În 5G, Managementul Serviciilor permit orchestrarea la nivelul

întregii rețele, atât pentru servicii, cât și pentru infrastructuri, asigurând implementarea dinamică a funcțiilor și programarea automată a conectivității între componentele rețelei. Acest model înlocuiește modul tradițional de configurare a rețelei, în care se gestionau individual echipamentele și straturile rețelei, folosind în schimb instrumente de management și orchestrare într-un mod coordonat. Pentru implementările de tip cloud-native, s-a concluzionat că containerele reprezintă o etapă evolutivă a tehnologiilor de virtualizare, rulând microserviciile pe deasupra sistemului de operare (OS).

Arhitectura pe bază de microservicii în containere aduce o serie de beneficii, printre care:

- consum eficient de resurse,
- creare și implementare rapidă a serviciilor,
- reducerea timpului de pornire a aplicațiilor,
- costuri mai scăzute,
- scalabilitate,
- reziliență,
- portabilitate,
- independență față de sistemul de operare.

Implementarea serviciilor 5G în cadrul conceptului de arhitectură bazată pe servicii (SBA) necesită adoptarea unor instrumente de rețea moderne, care să permită transformarea evoluției implementării rețelelor, de la funcții de rețea dedicate, rulate pe infrastructură fizică (bare-metal), către funcții de rețea software, prin intermediul virtualizării

Această secțiune detaliază componentele sistemului 5G actual, care constă dintr-o rețea centrală 5G (5GC) și o rețea de acces radio 5G (NG-RAN), interconectate prin rețeaua de transport. Sunt prezentate funcțiile principale ale rețelei centrale 5G, inclusiv AMF (Access and Mobility Management Function), UPF (User Plane Function) și SMF (Session Management Function).

2.2 Evoluția arhitecturii de rețelei

Secțiunea explorează evoluția arhitecturii rețelelor de telecomunicații, bazându-se pe proiecte de cercetare europene precum VITAL-5G, ADROIT-6G și 6G-INTENSE. Este prezentată o arhitectură de referință 5G E2E care abordează neconcordanțele dintre infrastructurile Telco și servicii, permițând crearea dinamică a serviciilor și implementarea aplicațiilor. Arhitectura

include o Platformă de Servicii care acționează ca "creierul rețelei", oferind module pentru monitorizare, inventariere și gestionarea ciclului de viață al aplicațiilor.

2.3 Evoluția sistemului telecom

Secțiunea descrie tranziția de la rețelele fizice tradiționale la rețelele autonome bazate pe AI, identificând patru etape principale de evoluție, Figura 2-2.

- Telco 1.0 - Rețele Fizice (1990-2010): Infrastructură tradițională bazată pe hardware fizic dedicat, cu configurare manuală și capacitate fixă.
- Telco 2.0 - Rețele Virtualizate: Combinație între echipamente fizice și funcții de rețea virtualizate, oferind flexibilitate și scalabilitate îmbunătățite.
- Telco 3.0 - Rețele Native Cloud: Implementare containerizată și orchestrată a funcțiilor de rețea, utilizând arhitecturi de microservicii pentru scalabilitate și reziliență.
- Telco 4.0 - Rețele Autonome: Rețele complet autonome cu capabilități avansate de AI/ML, care pot opera cu intervenție umană minimă, oferind auto-optimizare, auto-vindecare și predicție.

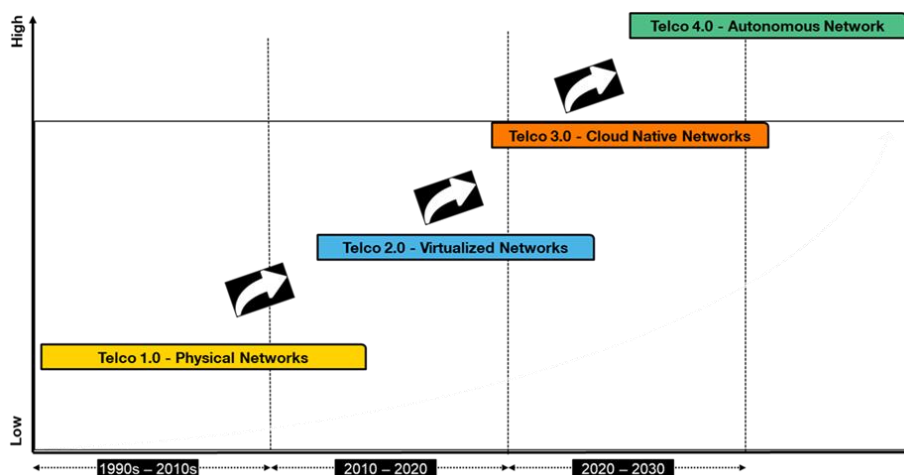


Figura 2-2 Evoluția sistemelor telecom

3. Funcționalități de administrare automată a rețelei

Capitolul explorează conceptul de " Funcționalități de management al rețelei fără intervenție umană" pentru viitoarele rețele 6G, prezentând o analiză aprofundată a abordărilor din industrie și cercetare privind automatizarea și managementul autonom al rețelelor.

3.1 Abordări standardizate în industrie

Analiza celor mai comune abordări din industrie privind automatizarea și configurarea auto-service a rețelelor:

1. 3GPP - Oferă informații relevante pentru managementul și orchestrarea operațiunilor, asigurarea și automatizarea, inclusiv interacțiunea cu entități externe. Specificațiile introduc conceptul de "Intent" (intenție) ca specificație formală a așteptărilor, cerințelor și constrângerilor date unui sistem tehnic.
2. ETSI ZSM - Furnizează conceptul de management zero-touch pentru rețele și servicii, concentrându-se pe principiile și conceptele bazate pe intenții pentru orchestrarea serviciilor. Specifică automatizarea în buclă închisă și aplicarea monitorizării, analizei și deciziilor bazate pe Machine Learning.
3. TM Forum - Prezintă o platformă colaborativă pentru profesioniștii din industrie, concentrată pe cadrul Open Digital Architecture (ODA), operațiuni autonome de rețea și experiențe client fără probleme. Ghidează transformarea digitală pentru noi tehnologii și servicii.

Accentul este pus pe 6G conceput și realizat ca un ecosistem inovator bazat pe servicii, tehnici de AI și ML, precum și adoptarea acestora de către industriile de telecomunicații. Acest lucru aduce o perturbare majoră pentru rețele, impactând proiectarea, implementarea procesării datelor, arhitectura inovatoare de rețea și abordarea viitoare a managementului. Așteptarea de la 6G este să furnizeze o infrastructură de înaltă performanță și eficientă energetic, unde serviciile de generație următoare pot fi dezvoltate și implementate. 6G este planificat pentru rețele cu inteligență bazată pe AI și abordare zero-touch. Există cercetări în industrie pentru operațiuni bazate pe software, zero-touch și, în cele din urmă, automatizarea fiecărui aspect al rețelei și al

serviciilor: specificațiile 3GPP¹ sau TM Forum² cu un program numit Proiectul Rețelelor Autonome³.

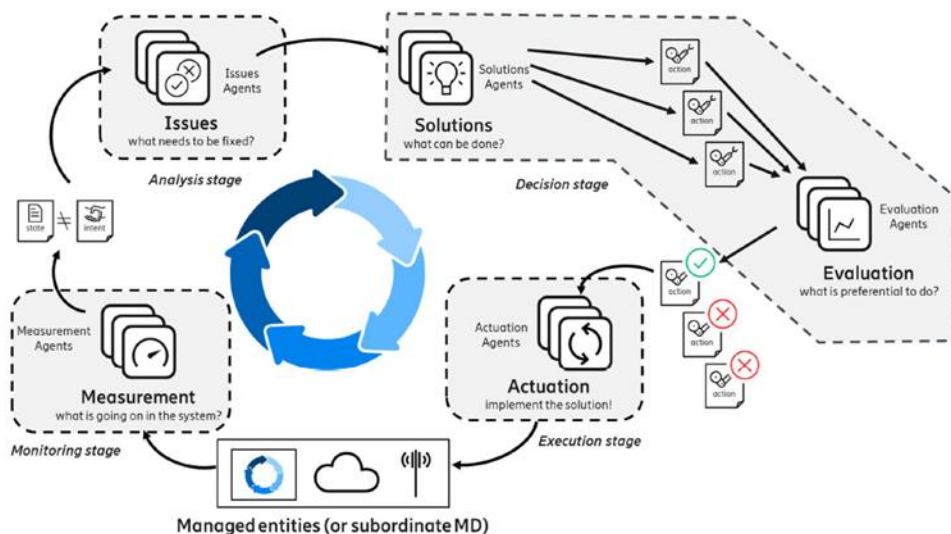


Figura 3-1 Buclă închisă ZSM[2]

Este esențial să avem un model de rețea care permite adaptări din mers, în special în mediul nativ cloud. Această arhitectură nouă, Figura 3-1, oferă fiabilitate și securitate, este bazată pe software și orientată spre operațiuni fără intervenție umană, deoarece reprezintă un mediu programabil consistent și de încredere. Operațiunile bazate pe AI/ML sunt facilitatori esențiali pentru scalabilitate și implementări la scară largă, iar abordarea zero-touch deschide operațiunile de management a resurselor pentru infrastructuri 6G auto-gestionate. Operațiunile sunt legate de soluțiile de management și control pentru arhitectura de management distribuită bazată pe servicii. Acest lucru este posibil prin bucle închise și automatizare bazată pe AI a rețelelor în diferite momente, prin coordonarea reacțiilor de control în timp real cu decizii de optimizare.

În industrie, conceptul ZSM⁴ utilizează paradigma Intenției, așa cum se arată în Figura 3-1, care rezumă abordarea bazată pe intenții:

¹ <https://www.3gpp.org/about-us>

² <https://www.tmforum.org/>

³ <https://www.tmforum.org/autonomous-networks-project/>

⁴ <https://www.etsi.org/technologies/zero-touch-network-service-management>

-
- Operațiunea bazată pe intenții transformă modelele tradiționale de interacțiune prin concentrarea pe cerințe mai degrabă decât pe acțiuni și procese explicite.
 - Sistemele care primesc intenții determină în mod autonom acțiunile și procesele adecvate necesare pentru îndeplinirea cerințelor declarate.
 - Modelarea formală și cadrele semantice comune sunt esențiale pentru o înțelegere clară între sisteme.
 - Limbajul natural poate servi ca metodă de introducere a datelor, necesitând interpretare contextuală și conversie la modelul standardizat de intenție.
 - Proprietarul intenției gestionează activ întregul ciclu de viață al fiecărei intenții.

3.2 Proiecte de cercetare europene

Smart Networks and Services Joint Undertaking (SNS JU) este un Parteneriat Public-Privat care își propune să faciliteze și să dezvolte leadership-ul industrial în Europa în domeniul rețelelor și serviciilor 5G și 6G. SNS JU finanțează proiecte care conturează o foaie de parcurs solidă pentru cercetare și inovare (C&I) și o agendă de implementare prin angajarea unei mase critice de părți interesate europene și facilitarea cooperării internaționale în cadrul diverselor inițiative 6G.

Capitolul continuă cu analiza proiectelor de cercetare europene SNS JU, în care autorul a fost implicat activ:

1. ADROIT-6G - Propune optimizări bazate pe AI/ML, software de rețea nativ în cloud și operațiuni software-driven, inclusiv ingineria arhitecturii. Introduce conceptul de "Crowdsourcing AI" și un cadru de management și orchestrare bazat pe AI.
2. 6G-INTENSE - Oferă o arhitectură nativă AI pentru inteligență bazată pe intenții, suportând abstractizarea compute-network. Introduce conceptul de Management și Orchestrare Distribuite Bazate pe Intenții (DIMO) pentru a aborda complexitățile rețelelor 6G.
3. VITAL-5G - Furnizează designul incipient, arhitectura, managementul, orchestrarea și automatizarea pentru transformarea rețelelor de telecomunicații către viitoarele rețele 6G.

3.3 Managementul slice-urilor de rețea

Capitolul prezintă conceptul de management al slice-urilor de rețea, explicând cum sistemele de rețea optimizează suportul pentru diverse servicii de comunicații. Sunt analizate două servicii principale: eMBB (Enhanced Mobile Broadband) și URLLC (Ultra-Reliable Low-Latency

Communications]), care pot fi utilizate pentru o varietate de aplicații, inclusive în implementarea lucrării.

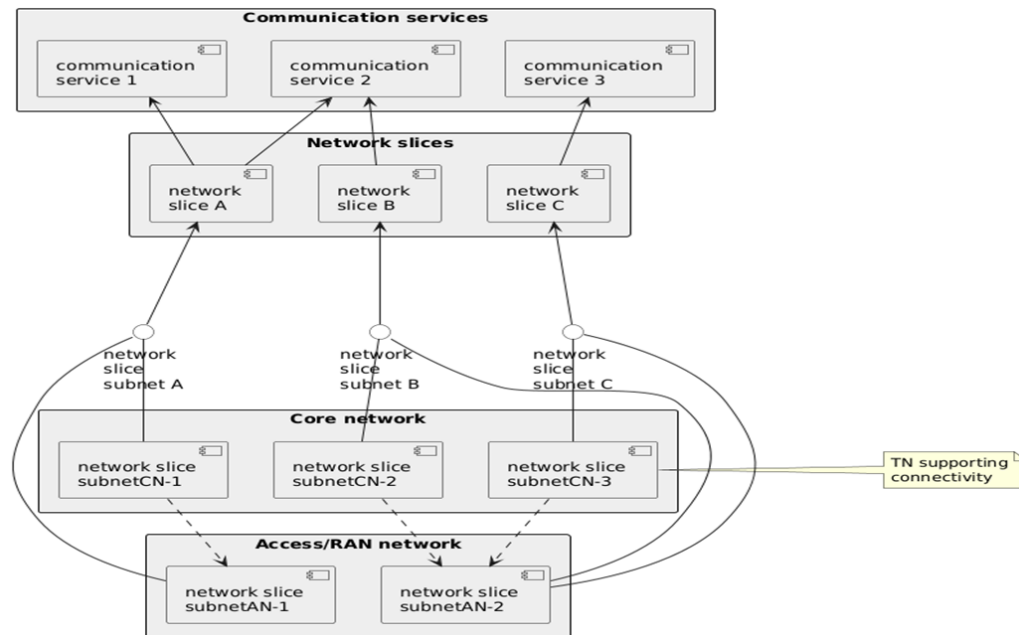


Figura 3-2 Servicii de comunicații furnizate de multiple slice-uri de rețea

Se pot considera în acest caz două servicii principale, descrise ca eMBB și URLLC, care pot fi utilizate pentru o varietate de aplicații. eMBB este capabil să ofere viteze și capacități de date mai mari pentru serviciile tipice de bandă largă mobilă, în timp ce Comunicațiile Ultra-Fiabile cu Latență Redusă (URLLC) prioritizează aplicațiile care necesită latență redusă și fiabilitate ridicată. Varietatea serviciilor de rețea impune cererea pentru slice-uri de rețea, urmărind aspecte legate de gestionarea serviciilor furnizate într-un mediu multi-slice, ca în Figura 3-2.

3.4 Arhitectura propusă pentru framework-ul AI/ML

Capitolul descrie o arhitectură de rețea inteligentă 6G cu integrare persuasivă AI/ML, construită pe baza 5G și având ca scop furnizarea de operațiuni autonome, utilizare optimizată a resurselor și agilitate îmbunătățită a serviciilor.

Arhitectura rețelei inteligente 6G reprezintă integrarea AI/ML, construită pe baza tehnologiei 5G și având ca scop furnizarea de operațiuni autonome, utilizarea optimizată a resurselor și agilitatea îmbunătățită a serviciilor. 6G introduce capacități native de AI/ML pentru inteligență avansată și are câteva componente ale nivelelor arhitecturale, aplicate unei infrastructuri cloud-native.

Arhitectura propusă include mai multe componente pe diferite niveluri:

1. Nivel de servicii: Manager bazat pe intenții, Manager de ciclu de viață al serviciilor, Manager de testare și experimentare.
2. Framework AI/ML: Model digitali de rețea, repozitor de modele AI/ML, Plan nativ AI, Framework AI distribuit.
3. Management și orchestrare: M&O distribuit bazat pe intenții, guvernanta în buclă închisă, Manager și orchestrator de domeniu/tenant.
4. Resurse: Management de slice-uri de rețea, Network Compute Fabric, Manager și orchestrator local.
5. Infrastructură: RAN, Transport, Core, Edge Computing

Această arhitectură unifică AI/ML în rețele pentru a realiza automatizarea zero-touch și performanțe îmbunătățite, oferind un management de rețea mai agil și mai eficient, cu coordonare nativă AI/ML.

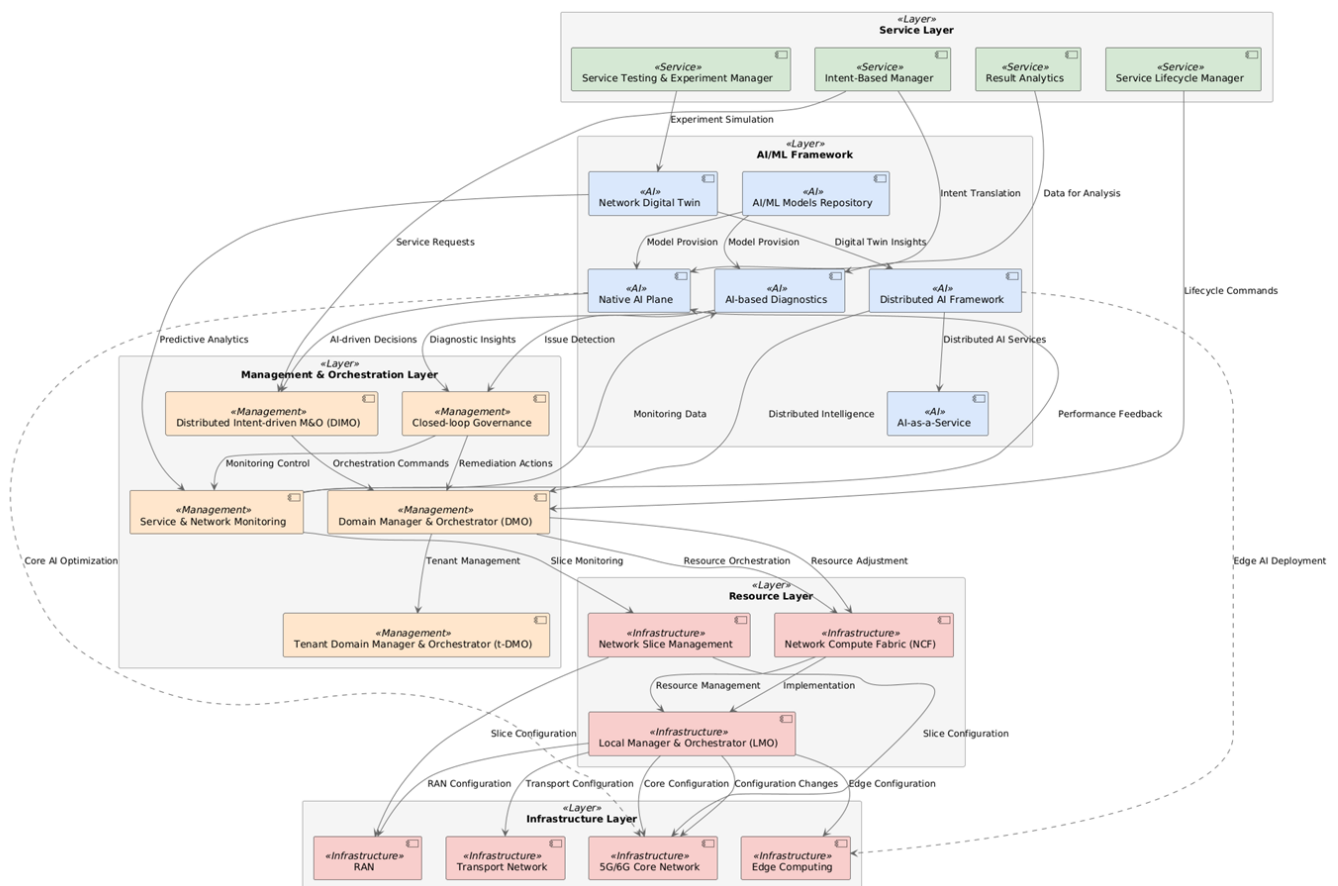


Figura 3-3 Propunere de arhitectură Telco cu suport AI/ML

Arhitectura extinde viziunea Zero Touch a managementului bazat pe intenții, prin soluția integrată end-to-end propusă pentru operațiuni zero-touch. Abordarea bazată pe intenții se concentrează pe "ce" în loc de "cum" și permite automatizarea prin bucle închise. Valorificând sarcinile AI/ML, aceasta susține optimizarea direcționată a resurselor, detectarea anomaliilor și predicția serviciilor. O arhitectură stratificată încorporează AI/ML la toate nivelurile, începând cu stratul de Servicii, care oferă interfețe bazate pe intenții pentru cererile de servicii și managementul ciclului de viață. Framework-ul AI/ML include AI pentru inteligența centralizată, depozitele de modele și NDT pentru cazurile de simulare a rețelei. Orchestrarea bazată pe intenții, guvernanța în buclă închisă și sarcinile de management specifice domeniului sunt implementate prin planul de Management și Orchestrare, care interacționează cu planul de Resurse pentru configurarea slice-urilor și a altor resurse, fiind susținut de orchestrarea locală a domeniilor (LMO) în componentele de rețea RAN, Core sau Transport.

4. Tehnici si Algoritmi

Capitolul prezintă tehnicile și algoritmi de inteligență artificială și învățare automată (AI/ML) care pot fi utilizați pentru a îmbunătăți capabilitățile rețelelor de telecomunicații 6G, cu accent pe automatizarea și optimizarea managementului rețelei.

4.1 AI/ML intro in telecom systems

Secțiunea introduce rolul AI/ML în rețelele 6G, subliniind că, deși AI/ML nu este un concept nou în industrie, reprezintă o inovație reală pentru viitoarele rețele de telecomunicații. Este propusă o arhitectură generală de rețea care integrează AI/ML pentru automatizare și optimizare, structurată în jurul conceptului de "as-a-Service" (aaS).

Conceptele AI asistă în creșterea capabilităților rețelei, noi configurații și algoritmi bazați pe modele de optimizare pentru configurarea rețelei și îmbunătățirea eficienței. Implementările și utilizarea 6G vor fi abordate cu AI și alocarea predictivă a resurselor bazată pe ML, care se concentrează pe provocările legate de problemele de performanță a rețelei. Există necesitatea de a introduce concepte generale legate de tehnicile AI/ML, comparația cu abordarea clasică de calcul, ce poate face AI/ML și care sunt prerechizitele care fac posibilă evoluția AI. Contextul algoritmilor AI/ML este de a utiliza eficient resursele, menținând în același timp performanța serviciilor și parametrii QoS în limitele intenției serviciului furnizat.

Arhitectura propusă permite:

- Implementarea automată a rețelei prin Network Digital Twin (NDT) as a Service
- Optimizarea QoS prin modele AI/ML antrenate
- Optimizarea resurselor pentru diferite domenii de rețea (RAN, Core, Edge)
- Adaptare în timp real prin sisteme cu buclă închisă
- Eficiență energetică și sustenabilitate

4.2 Premise necesare pentru realizarea AI/ML

Există mai multe atribute care pot face posibilă ML/AI, bazate pe următoarele caracteristici: Putere de calcul masivă, disponibilitatea unor seturi de date uriașe, algoritmi îmbunătățiți de învățare automată, cod open-source și framework-uri.

Seturile de date masive se referă la volume mari de date structurate sau nestructurate care sunt utilizate pentru a antrena modele și a îmbunătăți performanța acestora. Datele sunt cruciale

pentru avansarea tehnologiilor ML și AI, permițând modele mai precise și mai eficiente în diverse aplicații.

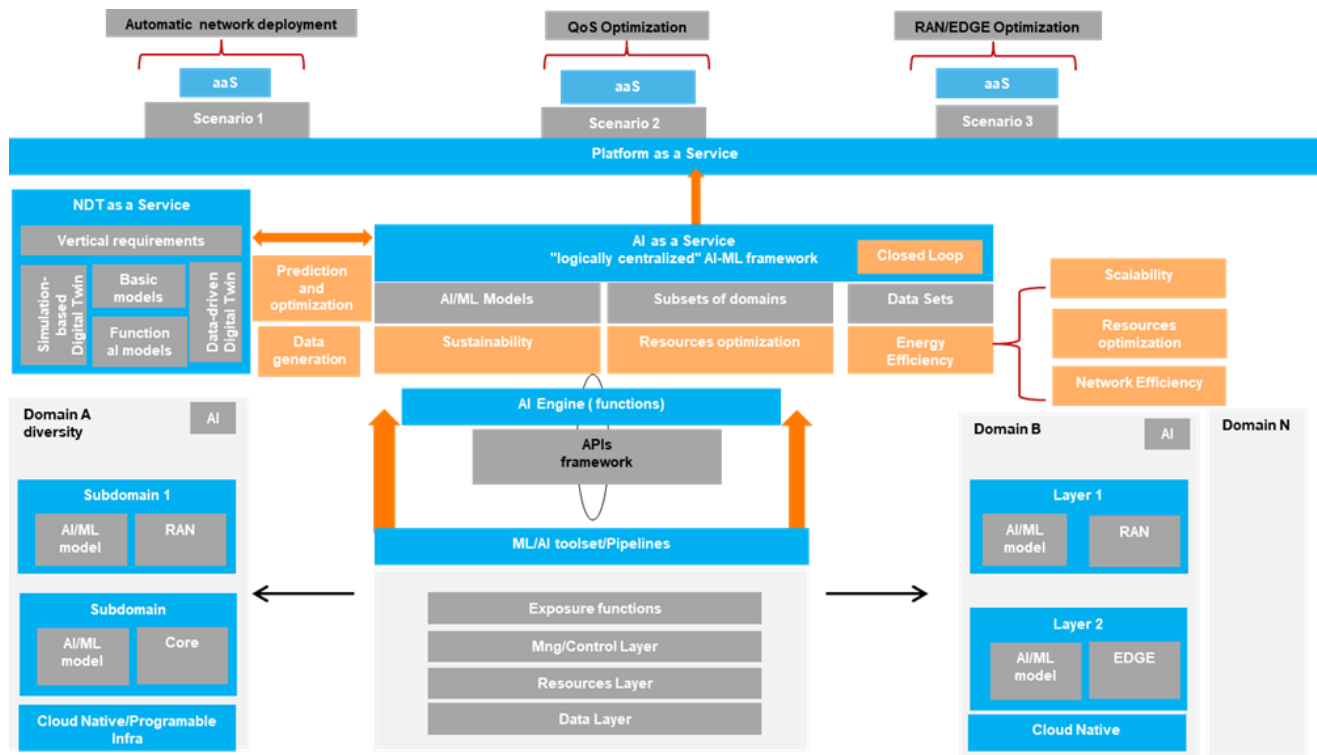


Figura 4-1 Prezentare generală a arhitecturii de rețea - cu accent pe AI/ML

Figura 4-1 prezintă evoluția spre 6G cu suport AI/ML a rețelelor. Aceasta abordează multiple scenarii și introduce conceptul de Network Digital Twin (NDT) ca Serviciu pentru simularea comportamentului rețelei. Arhitectura propune optimizarea QoS prin modele AI/ML și o logică centralizată cu bucle închise pentru ajustări automate. Soluția poate fi aplicată diverselor elemente de rețea (RAN, Core, Edge) și asigură scalabilitate. Pentru fiecare domeniu de rețea există integrări AI/ML specifice: modele pentru optimizarea performanței RAN, îmbunătățirea funcțiilor de rețea Core și infrastructură cloud nativă programabilă.

Arhitectura include definiții pentru modele AI/ML orientate spre sustenabilitate, optimizarea resurselor, eficiență energetică și predicție, pentru a satisface cerințele serviciilor cu costuri minime.

Algoritmi îmbunătățiți de învățare automata analizați:

Gradient Boosting Machines (GBM)

- Cum funcționează: GBM construiește modele secvențial, unde fiecare model nou corectează erorile făcute de cele anterioare. Combină învățători slabi (adesea arbori de decizie) pentru a crea un model predictiv puternic.
- Beneficii: Precizie ridicată, gestionează diverse tipuri de date și este robust la supraajustare cu reglajul corespunzător.
- Cazuri de utilizare: Date structurate
- Puncte forte: Precizie ridicată, flexibilitate și robustețe.

XGBoost (Extreme Gradient Boosting)

- Cum funcționează: O versiune optimizată a gradient boosting care include regularizare pentru a preveni supraajustarea, procesare paralelă pentru calcul mai rapid și eliminarea ramurilor arborilor pentru eficiență.
- Beneficii: Adesea depășește alte algoritme în competiții, foarte eficient și scalabil.
- Cazuri de utilizare: Date structurate/tabulare, competiții (de ex., Kaggle).
- Puncte forte: Precizie ridicată, flexibilitate și robustețe.

Random Forest

- Cum funcționează: O metodă de ansamblu care construiește mai mulți arbori de decizie în timpul antrenării și produce modul predicțiilor lor. Reduce supraajustarea prin medierea rezultatelor.
- Beneficii: Gestionează bine seturi mari de date, oferă importanța caracteristicilor și este robust la zgomot.
- Cazuri de utilizare: Sarcini generale de clasificare și regresie.
- Puncte forte: Robustețe la supraajustare și ușurință în interpretare.

Rețele Neuronale Convoluționale (CNN)

- Cum funcționează: Utilizate în principal pentru date de imagine, CNN-urile folosesc straturi convoluționale pentru a detecta automat caracteristici și modele în imagini. Ele reduc necesitatea extragerii manuale a caracteristicilor.
- Beneficii: Excelente pentru clasificarea imaginilor, detectarea obiectelor și alte sarcini vizuale.

-
- Cazuri de utilizare: Clasificarea imaginilor, detectarea obiectelor și sarcini de viziune computerizată.
 - Puncte forte: Capacități excelente de extragere a caracteristicilor pentru date vizuale.

Rețele Neuronale Recurente (RNN) și LSTM (Long Short-Term Memory)

- Cum funcționează: RNN-urile sunt concepute pentru date secvențiale, folosind bucle pentru a menține informații de-a lungul pașilor temporali. LSTM-urile, un tip de RNN, abordează problema gradientului care dispare, permițând învățarea dependențelor pe termen lung.
- Beneficii: Eficiente pentru predicția seriilor temporale, procesarea limbajului natural și recunoașterea vorbirii.
- Cazuri de utilizare: Deosebit de eficiente pentru sarcini care implică date secvențiale.
- Puncte forte: Gestionarea datelor secvențiale, făcându-le potrivite pentru o gamă largă de aplicații.

Transformers

- Cum funcționează: Utilizează mecanisme de auto-atenție pentru a cântări importanța diferitelor părți ale datelor de intrare, permițând procesarea paralelă și captarea dependențelor pe distanțe lungi.
- Beneficii: Performanță de ultimă generație în sarcini NLP, cum ar fi traducerea și generarea de text.
- Cazuri de utilizare: Procesarea limbajului natural (NLP), generarea de text și traducere.
- Puncte forte: Performanță de vârf în înțelegerea și generarea textului.

4.3 Soluție de achiziție a datelor

Colectarea datelor este efectuată folosind un sistem integrat de colectare a datelor, așa cum se vede în Figura 4-2, unde sistemul de sondare este conectat la o platformă centralizată cu o bază de date de serii temporale.

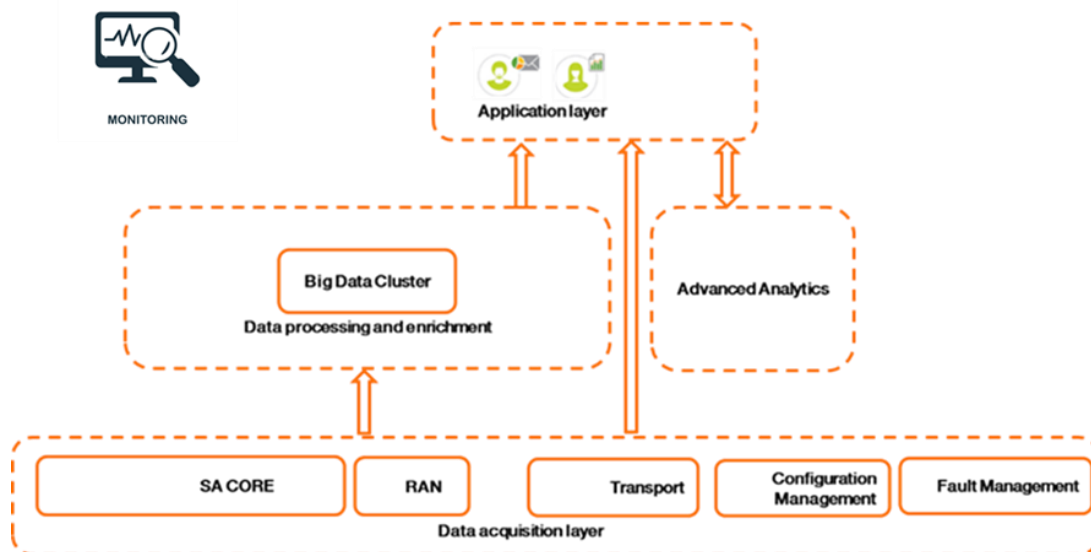


Figura 4-2 Soluție end-to-end pentru monitorizarea și achiziția datelor de rețea

În această diagramă se observă o arhitectură cuprinzătoare de monitorizare și analiză care este concepută pentru colectarea, procesarea și analizarea unor volume mari de date, ca date de serii temporale din infrastructura de rețea. Folosesc un sistem care urmează o abordare stratificată, capabil să gestioneze datele de la achiziție până la aplicații. În arhitectură, Figura 4-3, utilizez mai multe straturi: Nivelul de Achiziție a Datelor (Jos), Nivelul de Procesare a Datelor (Mijloc), Nivelul de Analiză Avansată (Mijloc Dreapta), Nivelul de Aplicație (Sus). Ca observație, această arhitectură de colectare a datelor și monitorizare a fost proiectată și implementată în Orange Romania 5GLAB⁵, aspectele de proiectare și arhitectură fiind realizate de autor. Această schemă propusă este deosebit de eficientă pentru rețelele de telecomunicații, deoarece orice sistem care generează volume mari de metrice cu marcat temporal care necesită analiză sunt mapate la această logică.

⁵ <https://5glab.orange.ro/en/>

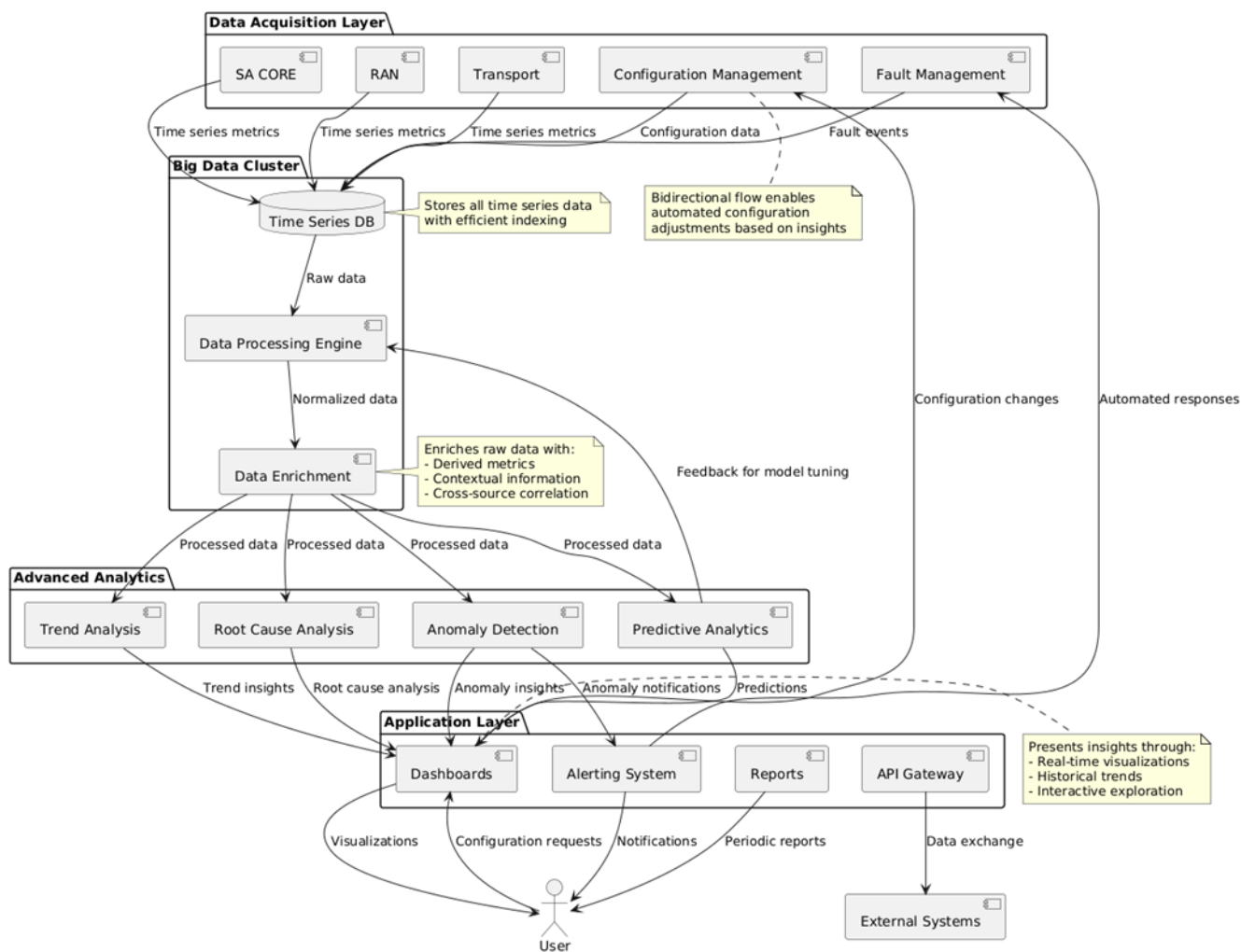


Figura 4-3 Modelarea arhitecturii de achiziție a datelor

4.4 Sistem de management al rețelei bazat pe inteligență artificială

Sistemul de management al rețelei bazat pe AI este o eficientizare a rețelei bazate pe AI, concentrându-se mai degrabă pe fluxurile de date și relațiile funcționale, extins ca arhitectură conceptuală decât pe componente individuale. Sistemul integrează capacități AI/ML pentru operațiuni inteligente de rețea, ca în Figura 4-4.

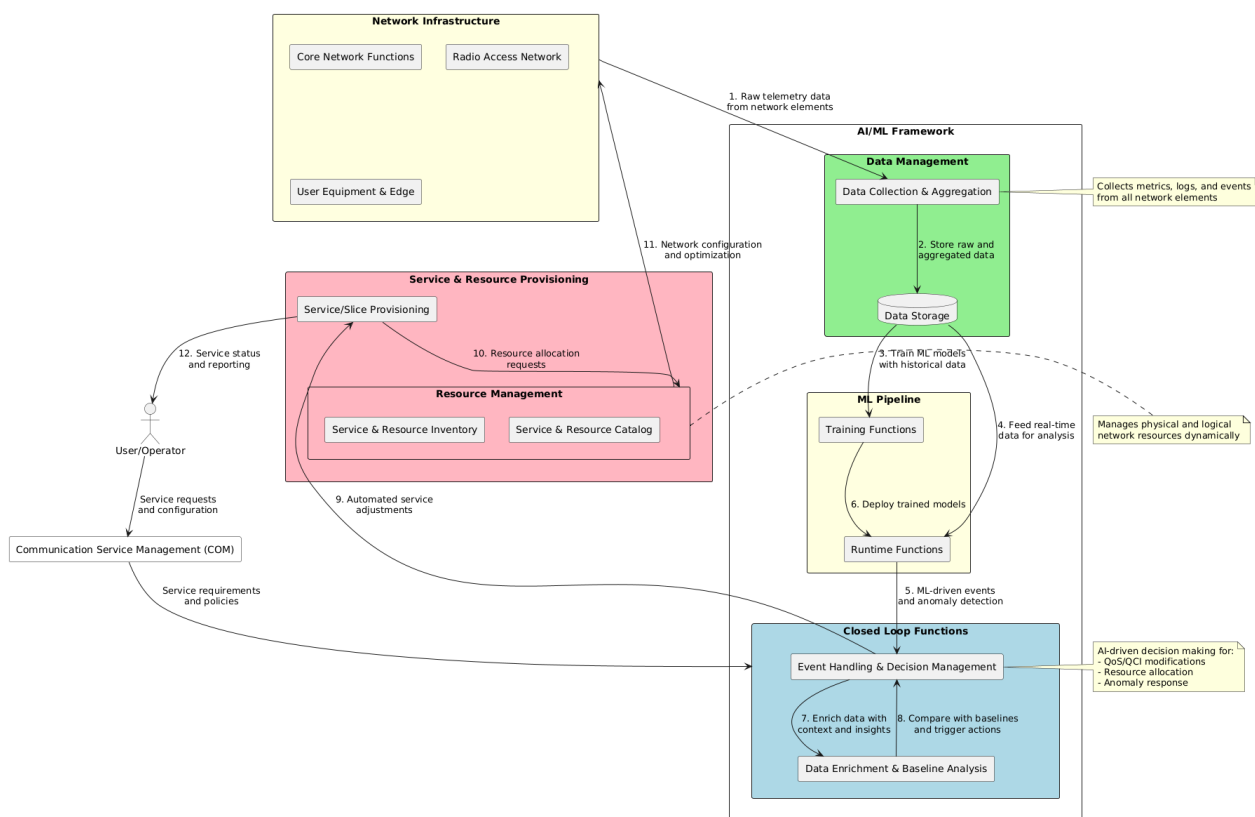


Figura 4-4 Fluxul sistemului de management al rețelei bazat pe inteligență artificială

- Faza de colectare a datelor (Pașii 1-2): acesta este punctul de plecare, procesul de colectare începe cu Infrastructura de Rețea (rețeaua centrală, RAN și edge) și monitorizează datele.
- Faza de procesare ML Pipeline (Pașii 3-6): aceasta este calea de antrenare, datele istorice colectate sunt utilizate pentru a antrena modele ML în componenta de funcții de antrenare.
- Faza de luare inteligentă a deciziilor (Pașii 7-8): aceasta este procesarea evenimentelor, componenta de gestionare a deciziilor care primește informații bazate pe ML și decide cu privire la detectarea anomaliilor.
- Faza de gestionare automată a serviciilor (Pașii 9-11): aceasta este compusă din ajustări automate ale serviciilor, bazate pe AI, pentru componenta de provizionare a Serviciului/Slice-ului.
- Faza de interacțiune cu utilizatorul (Pasul 12): acestea sunt cererile de servicii, prin inițierea procesului de solicitare a serviciilor sau efectuarea modificărilor de configurație.

4.5 Implementarea AI/ML și rezultatele inițiale

Având în vedere natura seriilor temporale a datelor colectate (utilizarea și alocarea resurselor în timp) și importanța captării dependențelor temporale, algoritmul Gated Recurrent Unit (GRU) pare a fi cel mai potrivit dintre cei patru algoritmi analizați.

Acesta înțelege în mod nativ secvențele și poate modela relațiile complexe dintre tiparele trecute și viitoare de utilizare a resurselor. Capacitatea sa de a reține memoria pe termen lung este esențială pentru captarea tendințelor, sezonality și a altor factori dependenți de timp care influențează alocarea resurselor. Diagrama de flux pentru antrenarea și testarea modelului GRU este prezentată în Figura 4-5, ilustrând procesul de dezvoltare a modelului de învățare automată pentru predicția performanței.

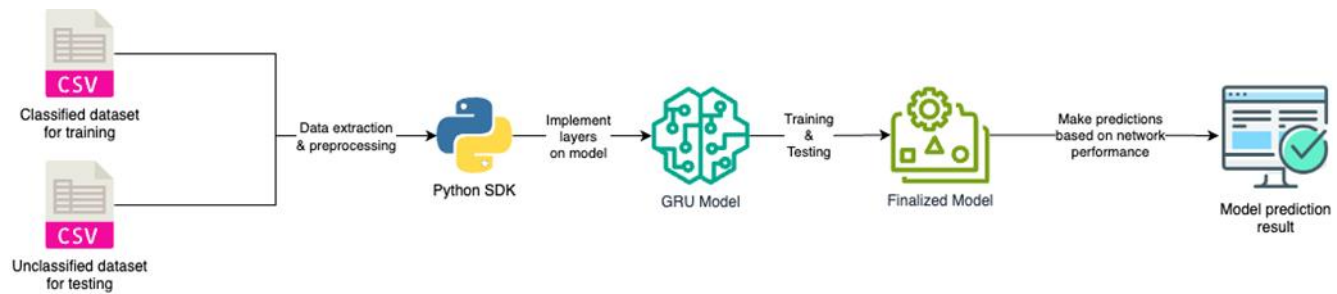


Figura 4-5 Modelarea GRU

În etapa Colectare date, sunt utilizate două seturi de date: unul clasificat pentru antrenarea modelului și unul neclasificat pentru testare.

În etapa de Procesare a datelor, seturile sunt supuse operațiunilor de extracție și preprocesare, urmate de modelarea datelor, pregătindu-le pentru antrenarea modelului.

Acestea sunt apoi transmise către un SDK Python pentru antrenarea efectivă a modelului.

În etapa de Testare, modelul antrenat este testat cu setul de date de test pentru a-i evalua performanța.

Sistemul de Învățare Automată în acest caz este capabil să genereze regulile care sunt încorporate într-un model antrenat. Sistemul învață din diverse exemple (date de antrenament), datorită faptului că nu este programat în mod explicit. Algoritmii selectați, combinați cu antrenamentul, creează regulile pe care le utilizează AI, iar modelul rezultat este capabil să rezolve sarcini. Când se furnizează date de intrare unei rețele neuronale, aceasta generează o predicție sau

o "estimare" despre intrare. Modelul antrenat rezultat este capabil să rezolve sarcini complexe în mod autonom. Antrenarea în "Învățarea Automată" reprezintă un model de învățare care este intensiv din punct de vedere computațional și necesită hardware specializat (GPU-uri) capabil să efectueze mii de operațiuni. Modelele de învățare automată sunt adesea simplificate și reduse (pruned) pentru a minimiza puterea de calcul și cerințele de memorie înainte de implementare, cu scopul de a optimiza performanța și a reduce consumul de resurse.

4.6 Seturi de date utilizate

Rețeaua este încărcată cu trafic injectat, pe două slice-uri de rețea dedicate: eMBB (Enhanced Mobile Broadband) și URLLC (Ultra-Reliable Low-Latency Communications), pentru care, prin intermediul automatizării, va fi implementată gestionarea dinamică a slice-urilor de rețea în timp real (live network slice management).

Această colecție de date este caracterizată prin traficul de tip downlink, care fluctuează semnificativ, indicând o încărcare variabilă a rețelei, și prin traficul de tip uplink, care rămâne relativ scăzut și stabil în comparație cu traficul downlink, pentru cele 2 tipuri de slice-uri.

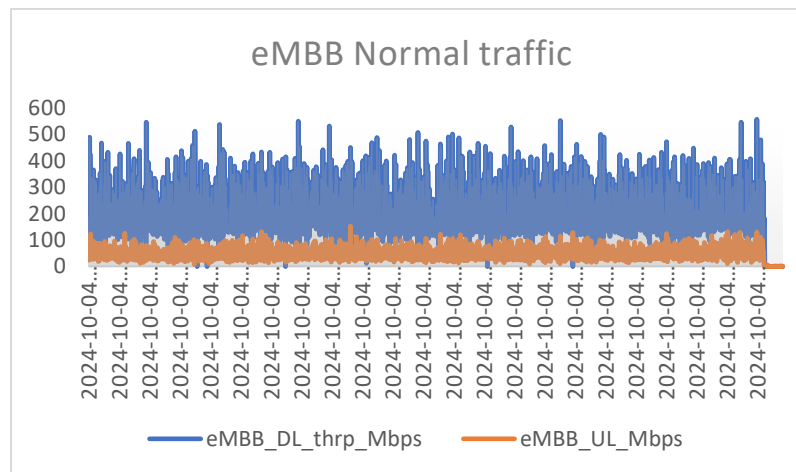


Figura 4-6 eMBB trafic normal

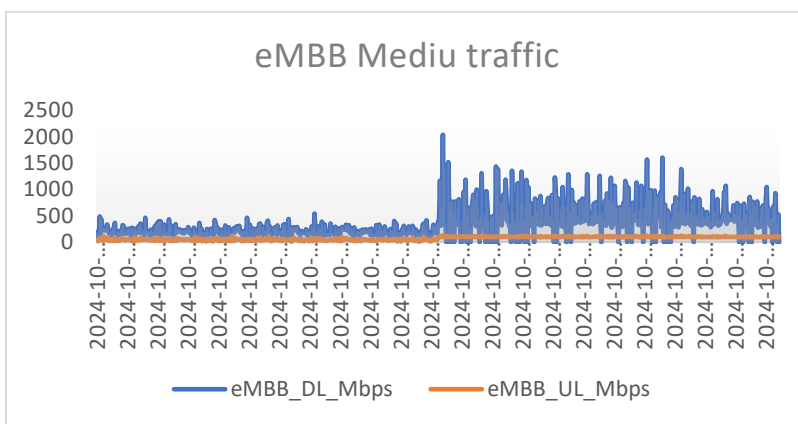


Figura 4-7 eMBB trafic mediu

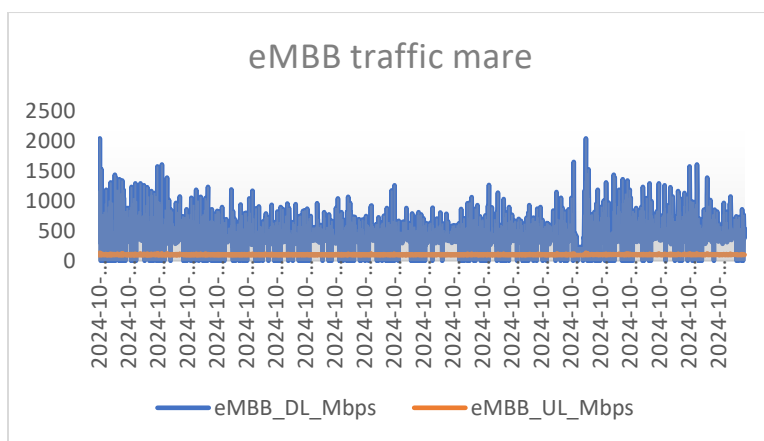


Figura 4-8 eMBB trafic mare

Graficele seriei temporale pentru eMBB High data traffic, prezentat în Figura 4-6, Figura 4-7, Figura 4-8, ilustrează un exemplu de date privind traficul de rețea. La intervale regulate de timp, acesta indică faptul că traficul de date colectat, exprimat în Mbps, variază până la pe diferite intervale de valori.

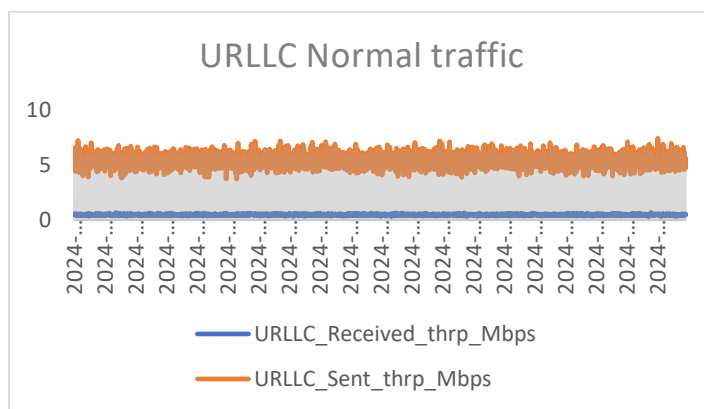


Figura 4-9 URLLC traffic normal

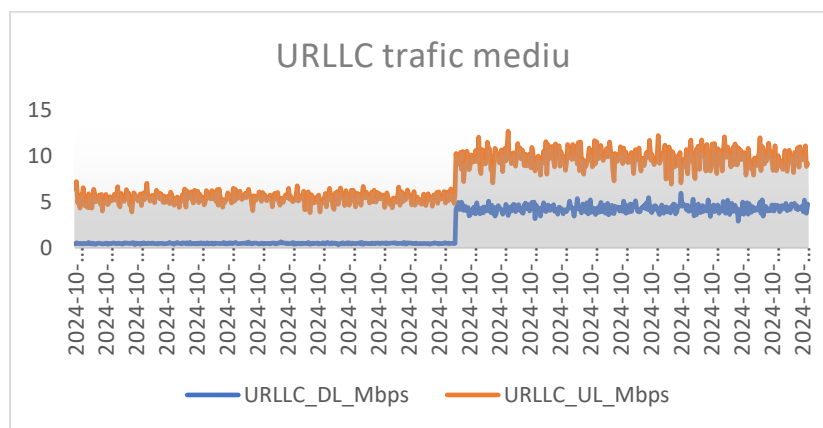


Figura 4-10 URLLC mediu trafic

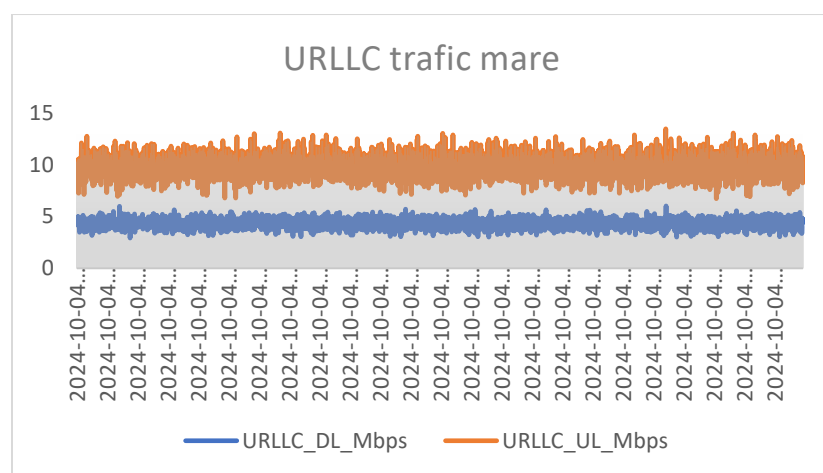


Figura 4-11 URLLC trafic mare

Graficele seriei temporale intitulat "URLLC" ilustrează exemple de date de trafic de rețea, în intervalele de timp regulate indică faptul că traficul de date variază până la 15Mbps.

Prin analiza seturilor de date și a traficului din grafice, în diverse scenarii, se poate observa că traficul unui serviciu de rețea poate influența traficul altora. Creșterea traficului poate conduce la congestionarea rețelei, afectând disponibilitatea benzii între diferite servicii, prin generarea de interferențe și perturbarea calității serviciului (QoS).

Traficul ridicat al unui serviciu poate genera interferențe sau poate intra în competiție cu alte servicii, în special în ceea ce privește resursele radio. Acest lucru afectează performanța serviciilor concurente care partajează aceeași infrastructură de rețea. Într-un scenariu real, rețeaua poate suferi degradări de performanță care afectează toate serviciile, inclusiv traficul de fundal sau serviciile cu prioritate redusă. Scopul acestei lucrări este de a reduce riscurile asociate acestor

situații prin utilizarea AI/ML și prin automatizarea proceselor, în vederea gestionării degradării serviciilor de rețea.

4.7 Dezvoltarea algoritmului

Pentru realizarea acestei sarcini sunt utilizate două seturi de date: unul pentru antrenare și unul pentru testare. Așa cum s-a menționat anterior, vor fi păstrate doar caracteristicile numerice, respectiv throughput-ul UL (uplink) și DL (downlink), exprimate în Mbps, pentru cele două categorii de servicii: eMBB (Enhanced Mobile Broadband) și URLLC (Ultra-Reliable Low-Latency Communications).

GRU este algoritmul selectat, este o formă specializată de Rețele Neuronale Recurente (RNN) concepută pentru a capta și valorifica dependențele temporale în date secvențiale. GRU-urile sunt deosebit de eficiente în scenarii în care ordinea și sincronizarea punctelor de date sunt cruciale, cum ar fi prognozarea seriilor temporale și procesarea limbajului natural, definit de formulele matematice:

$$\text{Update gate: } z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad 4.1$$

$$\text{Reset Gate: } r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad 4.2$$

$$\text{Candidate Activation: } \tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \quad 4.3$$

$$\text{Hidden State: } h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad 4.4$$

Pasul 1 reprezintă preprocesarea seturilor de date de antrenare și testare, inclusiv normalizarea (Figura 4-12). Datele sunt pregătite pentru antrenare și este evaluat un model de învățare automată prin selectarea caracteristicilor (de exemplu, throughput-ul exprimat în Mbps). Variabilele categoriale sunt codificate, variabila țintă este extrasă pentru setul de antrenare, iar caracteristicile sunt scalate. Flag-ul "is_training" controlează dacă variabila țintă și caracteristica „tip” sunt incluse, reflectând astfel diferențele dintre setul de antrenare și cel de testare.

Pasul 2 constă în definirea straturilor modelului, așa cum este ilustrat în Figura 4-13, utilizând modelul Gated Recurrent Unit (GRU) în Keras (o bibliotecă de rețele neuronale de nivel înalt scrisă în Python), pentru clasificare binară, de exemplu pentru predicția performanței rețelei (bună/slabă). Exemplul de cod definește și configurează o rețea neuronală recurentă (GRU) pentru a clasifica datele în două categorii (performanță bună/slabă). Straturile GRU gestionează date

secvențiale, ceea ce le face potrivite pentru serii temporale, așa cum sunt metricele de performanță ale rețelei colectate în timp.

Un obiect model Keras este creat, iar apoi poate fi antrenat folosind `model.fit()` cu datele preprocesate (`x_train, y_train`). După faza de antrenare, modelul poate fi utilizat pentru realizarea de predicții pe date noi, cu `model.predict()`.

Forma de intrare determină formatul datelor de input, specificând numărul de pași în timp (timesteps) și caracteristicile (features). Pentru un model GRU, inputul este de obicei tridimensional: numărul de eșantioane, timesteps și caracteristici. Parametrul `units` setează numărul de neuroni din fiecare strat GRU – un număr mai mare de unități permite modelului să captureze tipare mai complexe.

```
23 # Data Preprocessing
24 def preprocess_data(data, is_training=True):
25     # Extract features
26     features = ['BytesReceived', 'BytesSent', 'Received_thrp_Mbps', 'Sent_thrp_Mbps', 'URLLC']
27
28     if is_training:
29         features.append('Type') # Use 'Type' for training only
30
31     X = data[features]
32
33     # Encode the 'Type' column
34     if 'Type' in X.columns:
35         # Encode network volume (Type) as numerical: normal=0, normal-medium=1, medium=2, high=3
36         encoder = LabelEncoder()
37         X['Type'] = encoder.fit_transform(X['Type'])
38
39     if is_training:
40         y = data['Performance'] # Binary target (1 for good performance, 0 for poor performance)
41     else:
42         y = None
43
44     # Normalize features using MinMaxScaler
45     scaler = MinMaxScaler()
46     X_scaled = scaler.fit_transform(X)
47
48     return X_scaled, y
49
50 # Preprocess training and testing data
51 X_train, y_train = preprocess_data(training_data, is_training=True)
52 X_test, y_test = preprocess_data(testing_data, is_training=False)
53
```

Figura 4-12 Pre-procesarea datelor

```

58 # Build the GRU model
59 model = Sequential()
60 model.add(GRU(units=64, return_sequences=True, input_shape=(1, X_train.shape[2])))
61 model.add(Dropout(0.2)) # Dropout to prevent overfitting
62 model.add(GRU(units=64))
63 model.add(Dropout(0.2))
64 model.add(Dense(units=1, activation='sigmoid')) # Sigmoid for binary classification
65
66 # Compile the model
67 model.compile(
68     loss='binary_crossentropy',
69     optimizer=Adam(learning_rate=5.6e-5),
70     metrics=[BinaryAccuracy(), FalseNegatives()])
71
72

```

Figura 4-13 Definirea structurii modelului

Al treilea pas este antrenarea modelului (Model training), ilustrată în Figura 5.20. Antrenarea este controlată prin setarea dimensiunii batch-ului (batch size), care determină câte eşantioane sunt procesate înainte ca modelul să își actualizeze greutățile, și a numărului de epoci (epochs), care definește de câte ori modelul parcurge complet setul de date în timpul antrenării.

```

73 # Train the model
74 history = model.fit(X_train, y_train, epochs=20, batch_size=32, validation_split=0.2)
75

```

Figura 4-14 Antrenare model GRU

Procentul de testare este de 20%, iar restul de 80% din date este utilizat pentru procesul de antrenare. Codul antrenează modelul neuronal GRU definit anterior folosind datele preprocesate de antrenare, iar obiectul history va conține metricele de antrenare pentru fiecare epocă, inclusiv metricele de validare. Aceste metrice permit urmărirea modului în care modelul învață în timp, identificarea semnelor de overfitting, determinarea numărului optim de epoci și evaluarea performanței modelului pe datele care nu au fost văzute în faza de antrenare. Numărul de epoci este relevant, deoarece în învățarea automată o epocă reprezintă o parcurgere completă a setului de date de antrenare. În cazul de față, cu epochs=20, modelul va realiza 20 de parcurgeri complete ale datelor de antrenare, actualizându-și parametrii după fiecare batch în cadrul fiecărei rundă de antrenament. Epoca, ciclu de învățare reprezintă un ciclu complet de antrenare în care modelul vede toate datele de antrenament o singură dată.

5. Rezultate si Analiză

5.1 Mediul de testare

Activitatea de testare s-a desfășurat în cadrul Orange Romania 5GLAB, un mediu creat în scopul desfășurării de activități de cercetare, Figura 5-1.

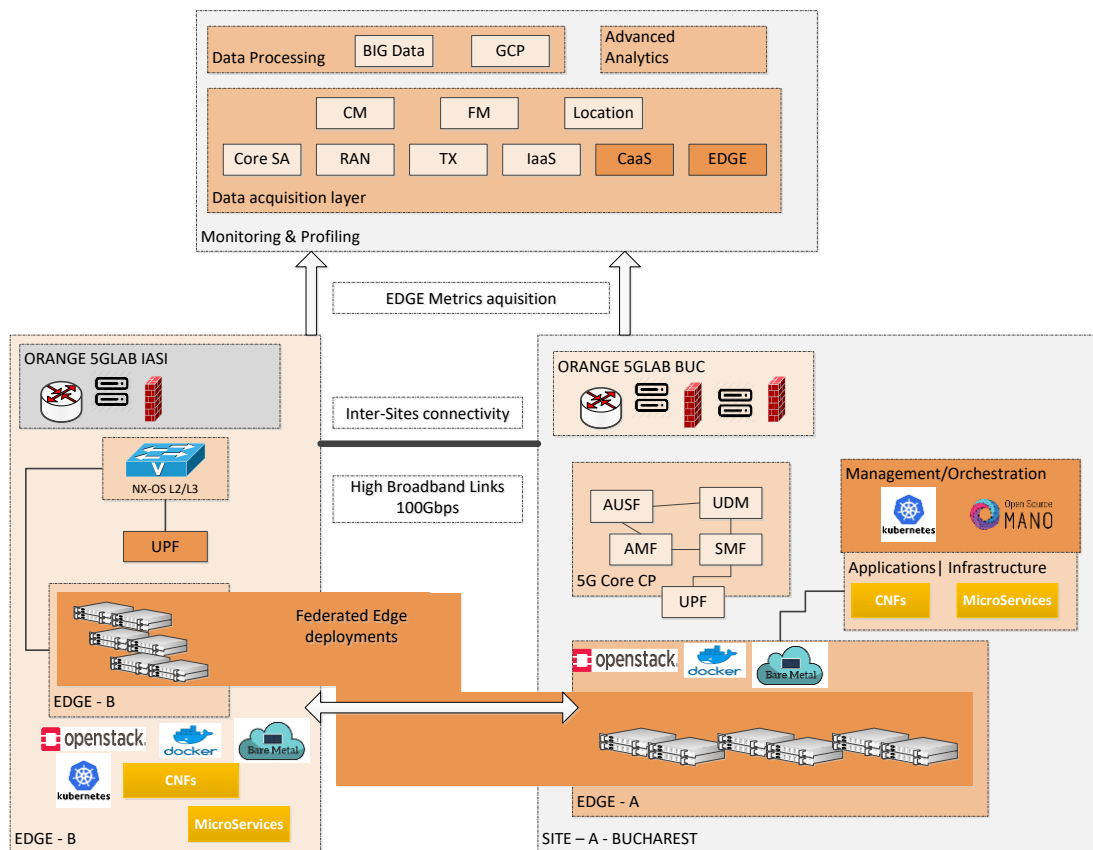


Figura 5-1 Mediul de experimentare 5GLAB

Scenariul de configurare al cazului de utilizare este asociat unui cadru AI/ML aplicat în rețea, destinat automatizării operațiunilor de servicii, prin predicția degradării unui serviciu de comunicații într-un slice de rețea și aplicarea de configurații de rețea pentru a restabili indicatorii de performanță (KPI) ai serviciului. Inteligența artificială și învățarea automată (AI/ML) vor permite operaționalizarea automată a serviciilor prin intermediul unui model manual, care realizează predicția degradării slice-ului de rețea, precum și printr-un mecanism AI de luare a deciziei pentru reconfigurarea slice-ului, în vederea restabilirii acordului de nivel de serviciu (SLA) în cazul unei degradări anticipat.

5.2 Scenariu de configurare

Scenariul de configurare a cazului de utilizare este legat de un cadru AI/ML de rețea pentru automatizarea operațiunilor de servicii în rețea, prin prezicerea degradării serviciului de comunicații într-un slice și aplicarea configurației de rețea pentru a restabili KPI-urile serviciului. AI/ML va permite operarea automatizată a serviciului printr-un model manual, care efectuează predicția degradării slice-ului de rețea.

A fost configurată serviciul de slice de rețea, activat end-to-end în toate elementele de rețea, cu un slice particular pentru un trafic de date (eMBB1), caracterizat prin Numele Slice-ului NSSAI-1, având Tipul de Serviciu Slice 1, diferențierea slice-ului 0xabcdef și DNN-ul suportat orange5glab.

Al doilea slice este caracterizat prin Numele Slice-ului NSSAI-2, având Tipul de Serviciu Slice 2, diferențierea slice-ului 0xabcdea și DNN-ul suportat orangeurlcc.

Rețeaua operatorului este împărțită în RAN, Core și Transport, unde slice-ul este configurat în întreaga rețea. Identificatorul slice-ului este identificarea slice-ului în fiecare parte a rețelei, permițând operatorului să aplice diferite politici și QoS fiecărui slice, și acest lucru asigură că fiecare serviciu va primi resursele și performanța de rețea așteptate și necesare.

5.3 Implementare unui model AI/ML

Prin intermediul stratului de orchestrare existent, un slice de rețea eMBB obișnuit este implementat în rețea, rolul orchestratorului fiind de a traduce cererea de serviciu într-o configurație de slice de rețea. După ce configurația inițială a slice-ului de rețea este stabilită, un utilizator poate accesa resursele de rețea. A doua configurație este de a implementa, folosind aceleași reguli ale orchestratorului, slice-ul de comunicație pentru un al doilea serviciu, care în acest caz poate fi considerat serviciul pentru care, în diverse condiții de încărcare, SLA-ul ar trebui să fie păstrat. În acest caz, în rețeaua 5G LAB, ambele servicii de rețea sunt configurate și monitorizate. Figura 5-2 descrie un cadru de orchestrare a rețelei bazat pe AI/ML cu planul de Orchestrare pentru servicii și resurse, prin gestionarea generală a furnizării serviciilor și alocarea resurselor de rețea. Ca observație, același plan de orchestrare va fi utilizat ulterior pentru a declanșa acțiuni în rețea, bazate pe predicții. În acest scenariu, guvernanta buclei închise controlează procesele automatizate de luare a deciziilor ca o soluție de management zero-touch.

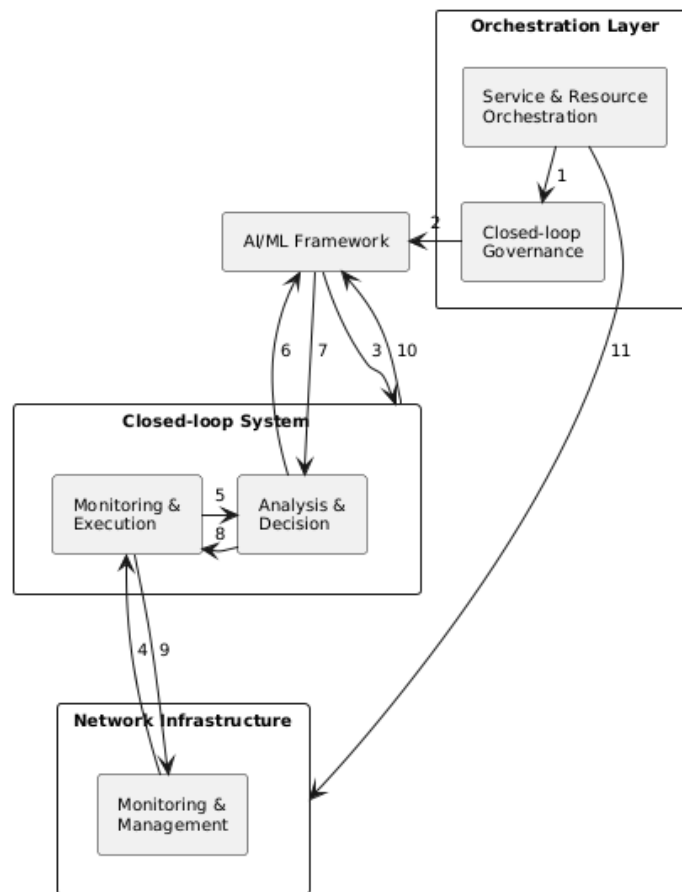


Figura 5-2 Soluție de implementare a slice-ului de rețea

Fluxul de configurare și implementare este inițiat (Pașii 1-3). După inițierea cererii de serviciu și implementarea modelului, sistemul de analiză va acționa ca o configurație de automatizare cu buclă închisă. Modelul este o componentă a blocului de Analiză și Decizie și, ca flux operațional, datele sunt colectate din rețea (metrici de performanță) și trimise către mediul de monitorizare și execuție (Pașii 4-5). Datele sunt formate și pregătite pentru analiză, cadrul AI/ML execută predicția și trimite în acest caz datele procesate pentru inferența modelului AI. Cadrul AI/ML returnează predicții (Pașii 6-7), bazate pe rezultatele modelului utilizat pe metricile actuale ale rețelei. În legătură cu AI, se trimite o acțiune către sistemul de management care declanșează modificări în rețea (Pașii 8-9). Modulul de execuție aplică modificările decise în infrastructura de rețea, modificările de configurație, alocarea resurselor și modificarea slice-ului. Sistemul propus este deschis pentru a oferi și bucla de feedback, unde sistemele cu buclă închisă pot trimite înapoi metrici de performanță legate de cadrul AI/ML pentru a îmbunătăți modelul AI și pentru a permite învățarea și adaptarea modelelor AI (Pasul 10). Feedback-ul de performanță nu face parte din lucrare și poate fi o activitate de lucru viitoare.

Exemplu de flux de trafic în timp real generat pentru slice-ul de rețea eMBB pentru DL (Downlink)

```
3983.01-3984.00 sec 44.6 MBytes 375.0 Mbits/sec  
[ 6] 3983.01-3984.00 sec 44.6 MBytes 375.0 Mbits/sec  
[ 8] 3983.01-3984.00 sec 44.6 MBytes 375.0 Mbits/sec  
[10] 3983.01-3984.00 sec 44.5 MBytes 374.0 Mbits/sec  
[SUM] 3983.01-3984.00 sec 178.3 MBytes 1499 Mbits/sec
```

Exemplu de flux de trafic în timp real generat pentru slice-ul de rețea eMBB pentru UL (Uplink)

```
[ 4] 3981.00-3982.01 sec 1.25 MBytes 10.4 Mbits/sec  
[ 6] 3981.00-3982.01 sec 1.25 MBytes 10.4 Mbits/sec  
[ 8] 3981.00-3982.01 sec 1.25 MBytes 10.4 Mbits/sec  
[10] 3981.00-3982.01 sec 1.25 MBytes 10.4 Mbits/sec  
[SUM] 3981.00-3982.01 sec 5.00 MBytes 41.6 Mbits/sec
```

Slice-ul de rețea de comunicații de interes este, de fapt, un trafic de date de analiză video care efectuează procesarea și identificarea fluxurilor în timp real. Pentru scopul cercetării și pentru a avea un mediu de testare realist, o aplicație care solicită un trafic garantat a fost instalată în mediul cloud-nativ, serviciul este implementat în rețea, iar aplicația de detecție începe să efectueze analize bazate pe configurația traficului video colectat în mediul de laborator. Slice-ul de rețea necesar este configurat prin intermediul Sistemului de Management al mediului de testare. La momentul implementării, modelul AI este activat și, pe baza măsurătorilor, efectuează predicții. Sistemul de monitorizare al mediului de testare achiziționează metrice din infrastructură, specifice pentru încărcarea traficului eMBB și pentru fluxul de analiză prioritarizat. Pe baza încărcării ridicate a traficului în slice-ul eMBB și datorită cererii de trafic de analiză pentru partea video, funcțiile de decizie declanșează modificări în rețea. Provocarea este de a antrena algoritmul pentru a prezice degradarea performanței serviciului și, de asemenea, de a impune în rețea modificările necesare pentru a restabili SLA-ul pentru serviciul efectuat.

5.4 Concluzii

Demonstrează o implementare practică a cadrului și activităților previzionate, în relație cu realizările zero-touch, utilizând tehnici și algoritmi, cu respectarea cadrului de implementare AI/ML. Configurația propusă este aplicată componentelor rețelei 5GLAB, folosind mecanismul deja existent pentru monitorizarea rețelei și execuția prin orchestrare.

Un prim punct al rezultatelor obținute este legat de arhitectura propusă și mai specific de componentele funcționale pentru elementele AI/ML integrate cu sistemul de management al rețelei din mediul de testare. Această implementare dovedește că în implementarea telco sunt

integrate diverse componente noi și, cu sprijinul AI, acestea pot fi operate pentru a asigura în viitor funcționalitățile serviciilor de rețea, așa cum este descris în teză.

Aceasta reprezintă unul dintre primele sisteme de management zero-touch pentru performanța serviciilor, aplicat sistemelor de management al slice-urilor de rețea cu o ajustare a QoS. Sistemul integrat este capabil să gestioneze alocarea resurselor și să mențină SLA-urile la nivelurile stabilite, în condițiile unui comportament variabil al comunicațiilor de rețea.

6. Concluzii finale, contribuții proprii și direcții de cercetare viitoare

Această teză investighează procesul de transformare a industriei de telecomunicații prin AI, de la tehnici ML/AI aplicabile pentru rețele de operare 6G cu zero touch, având ca scop automatizarea managementului rețelei prin orchestrare. Propune o arhitectură nouă pentru telecomunicații, cu un element central adăugat, AI/ML. Lucrarea prezintă viziunea viitoarelor rețele 6G, cu capacitatea zero-touch integrată în soluție. Prin explorarea AI/ML în telecomunicații, determinată de noile cazuri de utilizare emergente corelate și cu limitările sistemelor 5G actuale, lucrarea a propus o nouă arhitectură. Accentul este pus pe AI/ML, algoritmi potriviți, sistemul performant de achiziție de date, iar cercetarea descoperă potențialul AI/ML în telecomunicații în general și pentru a susține managementul automatizat și performanța auto-optimizantă a rețelei.

Elementele principale sunt, dincolo de AI/ML, conceptele de management bazat pe intenție, automatizarea cu buclă închisă, infrastructurile cloud-native, care împreună cu cadrele descrise și propuse oferă un model pentru dezvoltarea rețelelor inteligente. Evaluând algoritmi și modelele AI/ML, am identificat implementarea potrivită pentru o implementare particulară a predicției rețelei. Evaluând performanța, monitorizând astfel QoS-ul serviciului și metricile rețelei, este fezabil să se aplice analize pentru a ajusta setările rețelei pentru asigurarea SLA-urilor (aplicație video).

În același timp, este rezolvată problema automatizării în rețea, prin efectuarea dinamică a configurației de management a slice-urilor, demonstrează fezabilitatea și utilizabilitatea cadrului propus. Este ilustrat faptul că, într-o infrastructură reală de testare a rețelei, susținută de mediul de testare 5G al Orange România, AI/ML poate ajusta dinamic configurația rețelei, pentru a menține funcționalitatea unui serviciu în parametrii așteptați (asigurarea lățimii de bandă) în condiții de încărcare a rețelei. Este clar că există o legătură puternică între elementele rețelei, procesul ML/AI și acționarea și configurarea rețelei, în special în condiții de congestie a rețelei, fără nicio intervenție umană. Aceasta este o realizare, cazul de operare a rețelei zero-touch care validează cercetarea, evidențiază potențialul AI/ML în rețelele de telecomunicații și poate contribui la modelarea viitorului rețelelor de generație următoare, numite rețele 6G.

6.1 Contribuție proprie

Pentru a dezvolta cadrul arhitectural AI/ML propus, au fost realizați mai mulți pași, începând cu starea actuală a telecomunicațiilor și evoluția arhitecturii previzionate, pași incipienți

în domeniul rețelelor de management zero-touch pentru viitoarele rețele 6G. Am identificat problema și am propus pentru un subiect specific, managementul automat al rețelei, o soluție viabilă. Efortul a fost concentrat pentru a revizui și posibil a evolua cercetările existente actuale în 6G pentru rețele autonome. Acest lucru a fost realizat pe baza diverselor activități în desfășurare în domeniul 6G, în care sunt implicat, și prin încheierea cu propunerea unei noi arhitecturi 6G, AI/ML fiind considerată o componentă nativă.

Contribuția a evaluat, de asemenea, tehnicile și algoritmi existenți, ML/AI care pot fi aplicate în ecosistemul de telecomunicații și pentru a găsi soluții practice și implementări eficiente de cazuri de utilizare în contextul sistemelor de management al rețelei. Contribuția este legată și de mediul în care au fost efectuate activitățile, deoarece infrastructura rețelei 5G SA, cloud computing, sistemul de achiziție a datelor și procesarea datelor au fost construite sub coordonarea mea.

Principala contribuție este rezultatul cercetării, implementarea AI/ML în condiții specifice de rețea și aplicată practic, funcționalitatea end-to-end a arhitecturii propuse și configurația relevantă a rețelei pentru un scenariu de management zero-touch în conceptul de network slicing.

Rezumând contribuția, pot concluziona că au fost atinse mai multe obiective:

- Obiectivul 1, abordarea limitării actuale a rețelei, identificarea problemei și contribuția la soluție pentru un nou cadru arhitectural de rețea bazat pe AI/ML pentru a realiza managementul rețelei zero-touch în 6G. Acest obiectiv a fost tratat în contextul managementului bazat pe intenție, automatizării cu buclă închisă și a unei infrastructuri cloud-native, un nou design de cadru.
- Obiectivul 2, proiectarea arhitecturii de rețea bazată pe AI/ML, încorporând algoritmi ML pentru sarcinile de management al rețelei, asigurând predicția performanței rețelei și declanșarea ajustărilor QoS, cu GRU identificat ca fiind deosebit de potrivit.
- Obiectivul 3, avantajul disponibilității sistemului pentru colectarea și procesarea datelor de rețea din diverse surse, o demonstrație practică a achiziției de date, procesării datelor și analizei cu suport AI/ML este implementată într-un cadru pe un mediu de testare real. Studiul de caz analizează un management automatizat al slice-urilor de rețea și conformitatea SLA pentru o aplicație de analiză video. Această demonstrație validează funcționalitatea într-un mediu realist.

6.2 Lista Publicatiilor

1. M. Iordache, C. Patachia-Sultanoiu, O. Badita, B. Rusti, K. Trichias and J. Brenes, "5G Open Testbeds and Infrastructure extensions for NetApps advanced experiments", International Conference IEEE COMM 2022, Bucharest, June 2022
2. M. Iordache, O. Badita, B. Rusti, A. Bonea, G. Suci, E. Giannopolou, G. Landi, N. Slamnik-Krijestorac, "Future 5G network implementation and open testbeds deployment for real 5G experiments", 2022 IEEE Future Networks World Forum, Montreal 2022, doi:10.1109/FNWF55208.2022.00069
3. M. Iordache, R. Mihai, C. Patachia-Sultanoiu, J. Brenes, A. Ropodi, A. Margaritis, G. Suci, A. Vulpe, N. Slamnik-Kriještorac, A. Gavrielides, E. Gianopoulo, "Advanced 5G Open Testbed for Network Applications Experiments", Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 2023 July 26th, doi: 10.1109/EuCNC/6GSummit58263.2023.10188342, <https://ieeexplore.ieee.org/document/10188342>
4. M. Iordache, C. Patachia-Sultanoiu, "5G powered smart lighting in smart cities. SliceNet's smart city pilot" Eurescom Message, The magazine for telecom insiders, ICT for Smart Sustainable Cities, winter 2020, p25-26, <https://www.eurescom.eu/wp-content/uploads/2021/02/Eurescom-message-Winter-2020.pdf>
5. B. Sayadi, C. Chang, C. Tranoris, M. Iordache, K. Katsaros, Network Applications: Opening up 5G and beyond networks, V2.0, 5G-PPP Software Network Working Group, July 2023, DOI:10.5281/zenodo.7123919
6. 1.2.7 V. Charpentier et al., "Dynamic and Quality-aware Network Slice Management in 5G Testbeds," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 611-616, doi: 10.1109/EuCNC/6GSummit58263.2023.10188347.
7. 1.2.9 M. Iordache et al., "Advanced 5G Open Testbed for Network Applications Experiments," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 526-531, doi: 10.1109/EuCNC/6GSummit58263.2023.10188342.

-
8. 1E. Sender et al., „AI-Driven Zero-Touch Slice Optimization for 6G Networks”, 2024 IEEE 29th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), doi: DOI: 10.1109/CAMAD62243.2024.10942673
 9. White Paper, NETWORK AND SERVICE MANAGEMENT ADVANCEMENTS Key Frameworks and Interfaces towards Open, Intelligent and Reliable 6G networks, December 2024, DOI: 10.5281/zenodo.14234897
 10. A. Boutouchenet et al., „6G-INTENSE: Intent-Driven Native Artificial Intelligence Architecture Supporting Network-Compute Abstraction and Sensing at the Deep Edge”, Published in: IEEE Vehicular Technology Magazine; <https://ieeexplore.ieee.org/document/10844301>
 11. V. Charpentier et al., „Utilizing the VITAL-5G platform to advance 5G standalone integration with vertical industries”, EURASIP Journal on Wireless Communications and Networking; <https://jwcn-urasipjournals.springeropen.com/articles/10.1186/s13638-025-02456-x>

6.3 Direcții viitoare

Este evident că mai rămâne mult de discutat, deoarece dincolo de arhitectura propusă și cazul de validare, există încă întrebări deschise pentru diversele cazuri de utilizare și scenarii care pot fi analizate și în legătură cu modul în care industria va adopta și implementa astfel de cazuri și activități pentru managementul zero-touch.

Optimizarea rafinată a modelului AI/ML, explorarea altor modele și tehnici avansate de optimizare pentru a îmbunătăți precizia predicției și a reduce complexitatea computațională, inclusiv modele de învățare profundă și abordări de învățare prin consolidare.

Optimizarea multi-servicii/slice, pentru a extinde serviciul unic actual și slice-ul de rețea unic la un mediu multi-servicii și multi-slice, inclusiv dezvoltarea de noi modele AI/ML care pot oferi o astfel de optimizare.

Integrarea cu managementul bazat pe intenție, cadrul AI/ML propus să fie integrat cu sisteme bazate pe intenție, inclusiv dezvoltarea de mecanisme pentru a efectua traducerea de la intenții de nivel înalt în acțiuni și politici AI/ML specifice.

References

- [1] Chowdhury, Mostafa Zaman, et al. "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions." *IEEE Open Journal of the Communications Society* 1 (2020): 957-975.
- [2] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078*.
- [3] E. Ekudden. "Five Network Trends: Towards the 6G Era." *Ericsson Technology Review* 2021.9 (2021): 2-10.
- [4] R. Schwartz, et.al., "Green AI," *Communications of the ACM*, 2020
- [5] B.M. Khorsandi, M. Hoffmann, M. Uusitalo, M-H. Hamon, B. Richerzhagen, G. D'Aria, A. Gati, D. Lopez (editors), *Deliverable D1.3: Targets and requirements for 6G - initial E2E architecture*, Hexa-X project, 2022, available online <https://hexa-x.eu/wp-content/uploads/2022/03/Hexa-X D1.3.pdf>.
- [6] A. S. G. Andrae and T. Edler, "On global electricity usage of communication technology: Trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015. [Online]. Available: <http://www.mdpi.com/2078-1547/6/1/117>
- [7] P. Almasan, M Ferriol-Galmes, J. Paillisse, J. Suarez-Varela, "Network Digital Twin: Context, Enabling Technologies and Opportunities", *EEE Communications Magazine* 60(11):1-13, Nov 2022, DOI:10.1109/MCOM.001.2200012.
- [8] Saghezchi, F. B., Rodriguez, J., Vujicic, Z., Nascimento, A., Huq, K. M. S., & Gil-Castiñeira, F., "Drive towards 6G", In *Enabling 6G Mobile Networks* (pp. 3-35). Springer, Cham, 2022.
- [9] D. Scotece, C. Fiandrino, L. Foschini, "A Practical way to Handle Service Migration of ML-based Applications in Industrial Analytics," *IEEE GLOBECOM 2022*, 2022, doi: <https://doi.org/10.1109/GLOBECOM48099.2022.10001449>
- [10] B. Martini, M. Gharbaoui, P. Castoldi, "Intent-based network slicing for SDN vertical services with assurance: Context, design and preliminary experiments," *Future Generation Computer Systems*, Vol. 142, 2023