FEDGEN_HEALTHEDUAPP
(DEVELOPMENT OF CLOUD AND MOBILE HEALTH EDUCATION APPS FOR FEDGEN PLATFORM)

Presented by Prof. E. Adetiba and Dr. Itunnu Isewon

(CApIC-ACE)

@ The 1st DSTN Symposium, Gaston Berger University, Saint Louis, Senegal , 9th December., 2021.

THE WORLD BANK

AFD

IRD Institut de Recherche pour le Développement FRANCE

Inría

TEAM MEMBERS/FUNDERS

1. CApIC-ACE Team Members
   - ✓ Prof. E. F. Adebiyi - Co-Principal Investigator
   - ✓ Prof. E. Adetiba - Co-Principal Investigator
   - ✓ Dr. J. A. Badejo   - Co-Investigator
   - ✓ Dr. Isewon – Co-Investigator (CApIC-ACE Malaria Group)
   - ✓ Dr. Solomon Rotimi – Co-Investigator (CApIC-ACE Cancer Group)
   - ✓ Engr. Matthew B. Akanle – System Engineer
   - ✓ Engr. Dipo Olaleye - System Engineer
   - ✓ Mrs. Priscilla Ajayi – Software Developer

2. Digital Science and Technology Network (DSTN) Collaborators (CEA-MITIC) –
   - ✓ Prof. Maissa Mbaye (Co-PI)
   - ✓ Dr. Dame Diogoue (Co-Investigator).

3. Software Development Interns in CApIC-ACE and CEA-MITIC

4. Academic and Industry Partners –Landmark University, ABUAD, ULPGL DRC, IBM, HP, Inlak, Oracle.

5. Funding Agencies: AFD(ACEPartner), Covenant University, World Bank (ACE Impact) .
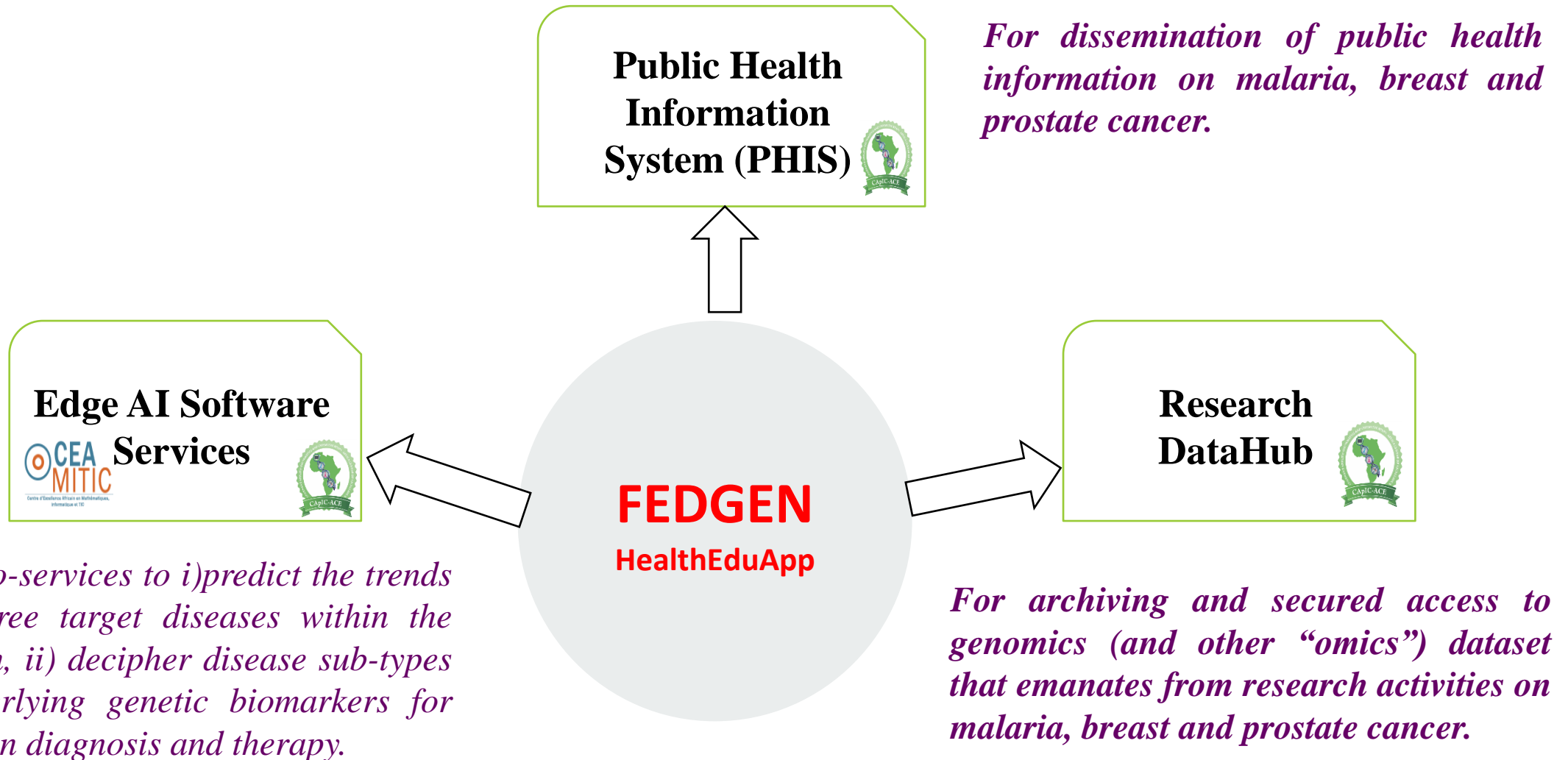
© CApIC-ACE

2

1.0   INTRODUCTION

- This is a Development Project being funded by AFD through the DSTN.

- The duration of the project is 2 years (2020-2022) and the components been funded are -

      i) Internship

      ii) Mobility

      iii) Operational Cost

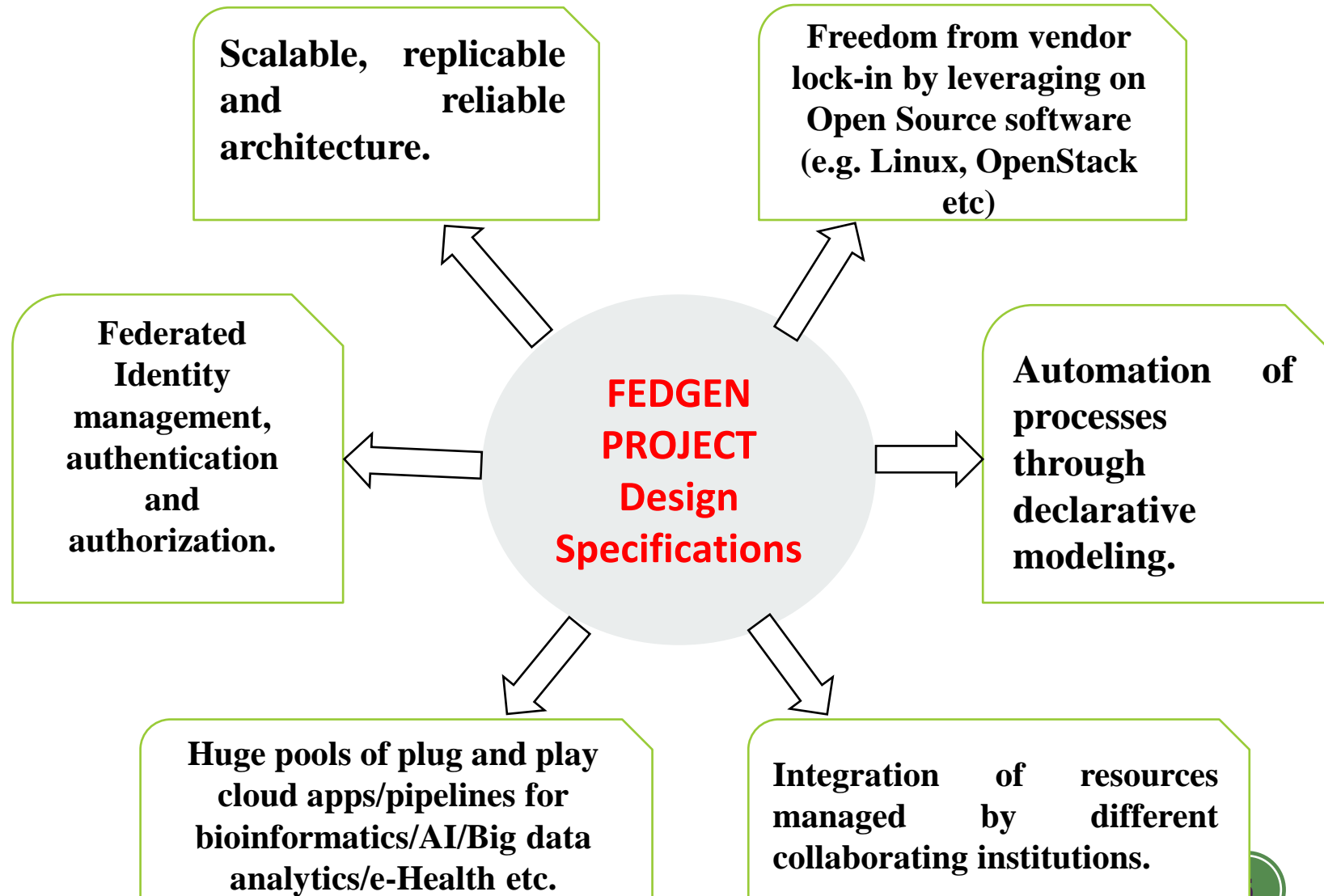# 2.0 PROJECT OBJECTIVES

Public Health Information System (PHIS)

For dissemination of public health information on malaria, breast and prostate cancer.

Edge AI Software Services

CEA MITIC

FEDGEN
HealthEduApp

Research DataHub

With micro-services to i)predict the trends of our three target diseases within the population, ii) decipher disease sub-types with underlying genetic biomarkers for precision in diagnosis and therapy.

For archiving and secured access to genomics (and other "omics") dataset that emanates from research activities on malaria, breast and prostate cancer.
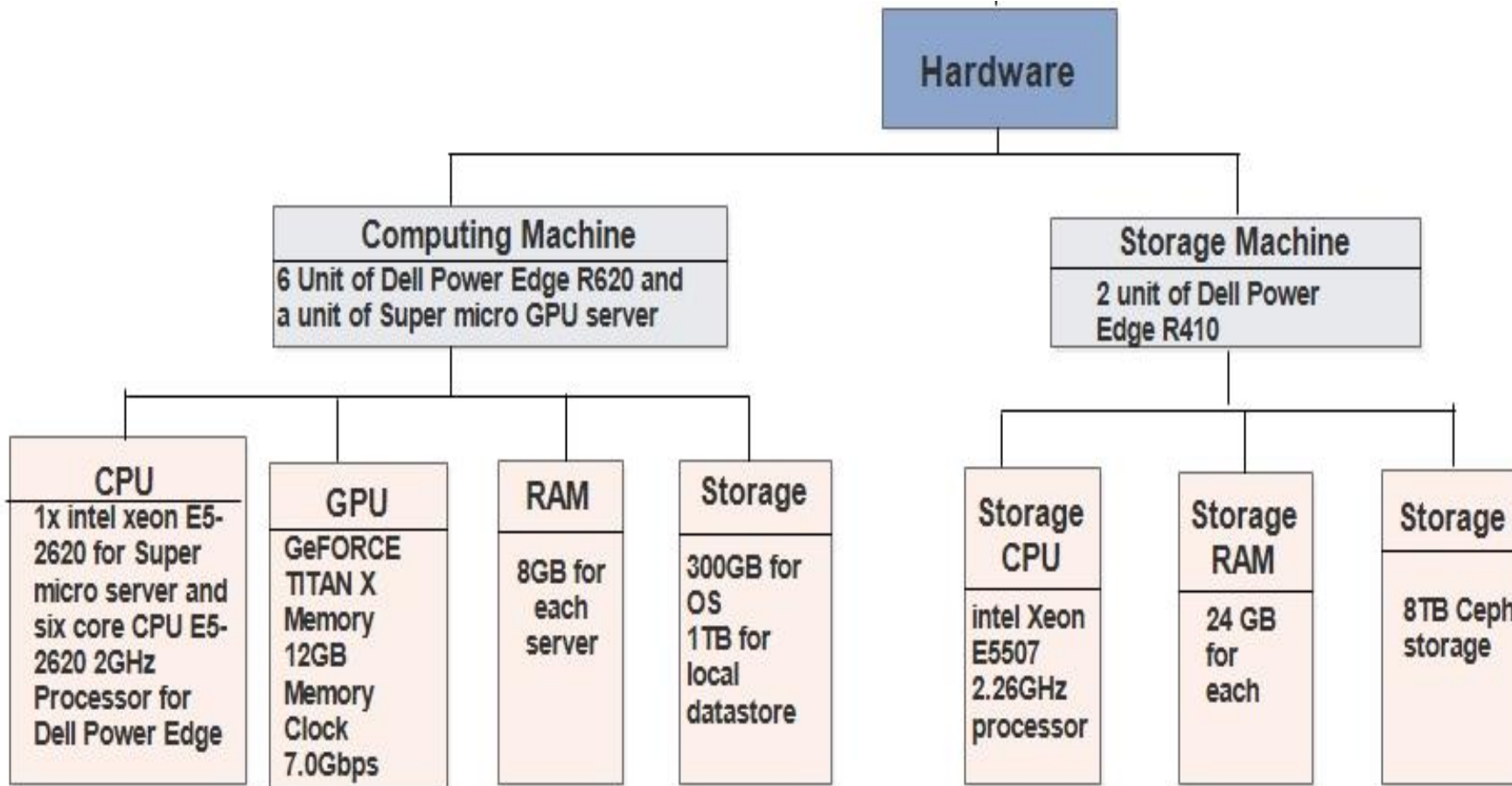
3.0 STATE OF PROGRESS

FEDGEN Testbed: The prototype deployment platform for our cloud app projects has been successfully setup.

(System Engrs., 2 CApIC-ACE M.Eng. Students, 1 Intern)

Scalable, replicable and reliable architecture.

Freedom from vendor lock-in by leveraging on Open Source software (e.g. Linux, OpenStack etc)

Federated Identity management, authentication and authorization.

FEDGEN PROJECT Design Specifications

Automation of processes through declarative modeling.

Huge pools of plug and play cloud apps/pipelines for bioinformatics/AI/Big data analytics/e-Health etc.

Integration of resources managed by different collaborating institutions.

- For the FEDGEN Testbed, we leveraged on open source software such as:
  - *Linux,*
  - *OpenStack,*
  - *MaaS,*
  - *Juju,*
  - *Juju Charms* etc.

- Publication Output: A conference paper presented @ ICIIA 2021 to be published in a Springer Proceeding (Scopus Indexed).
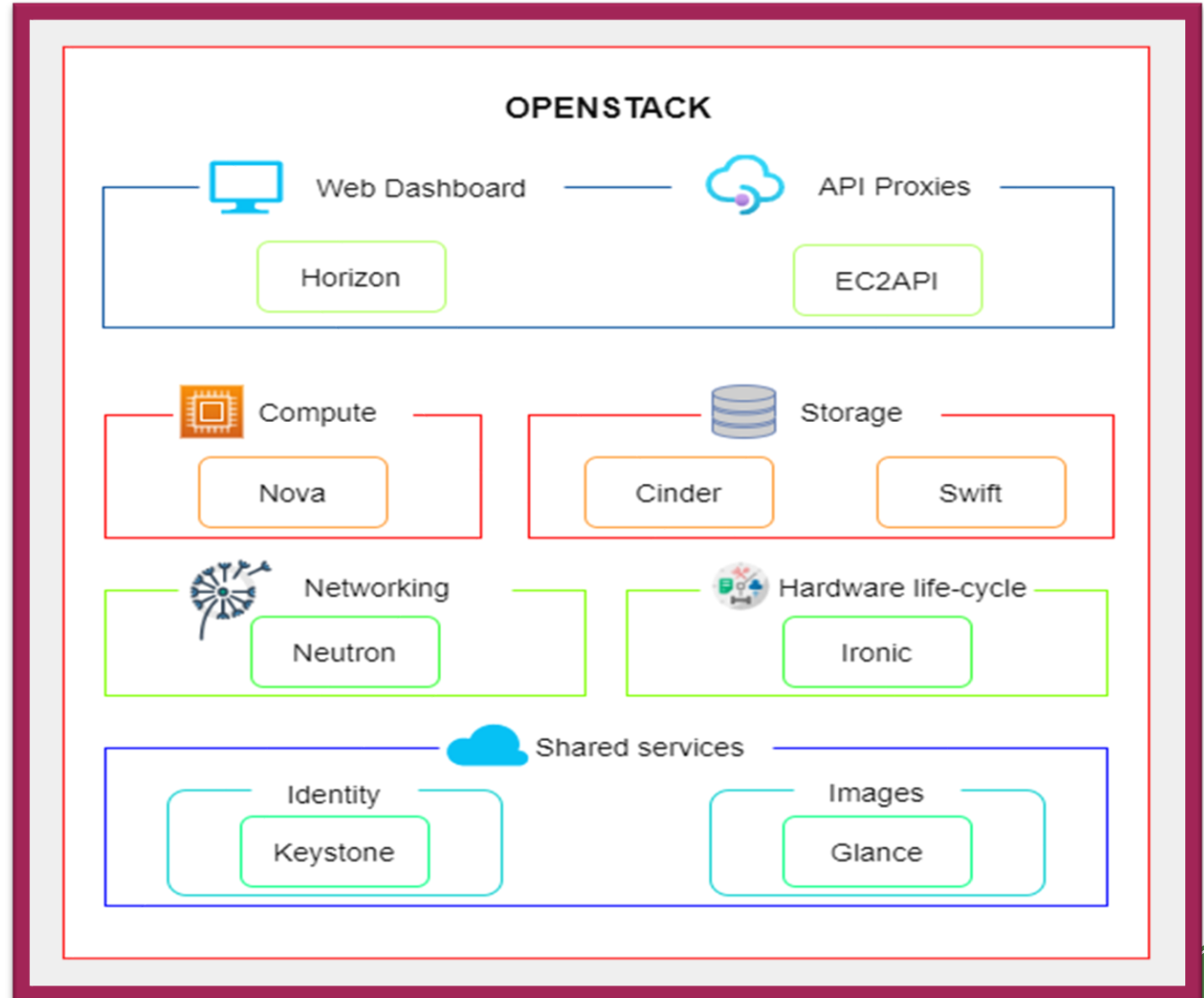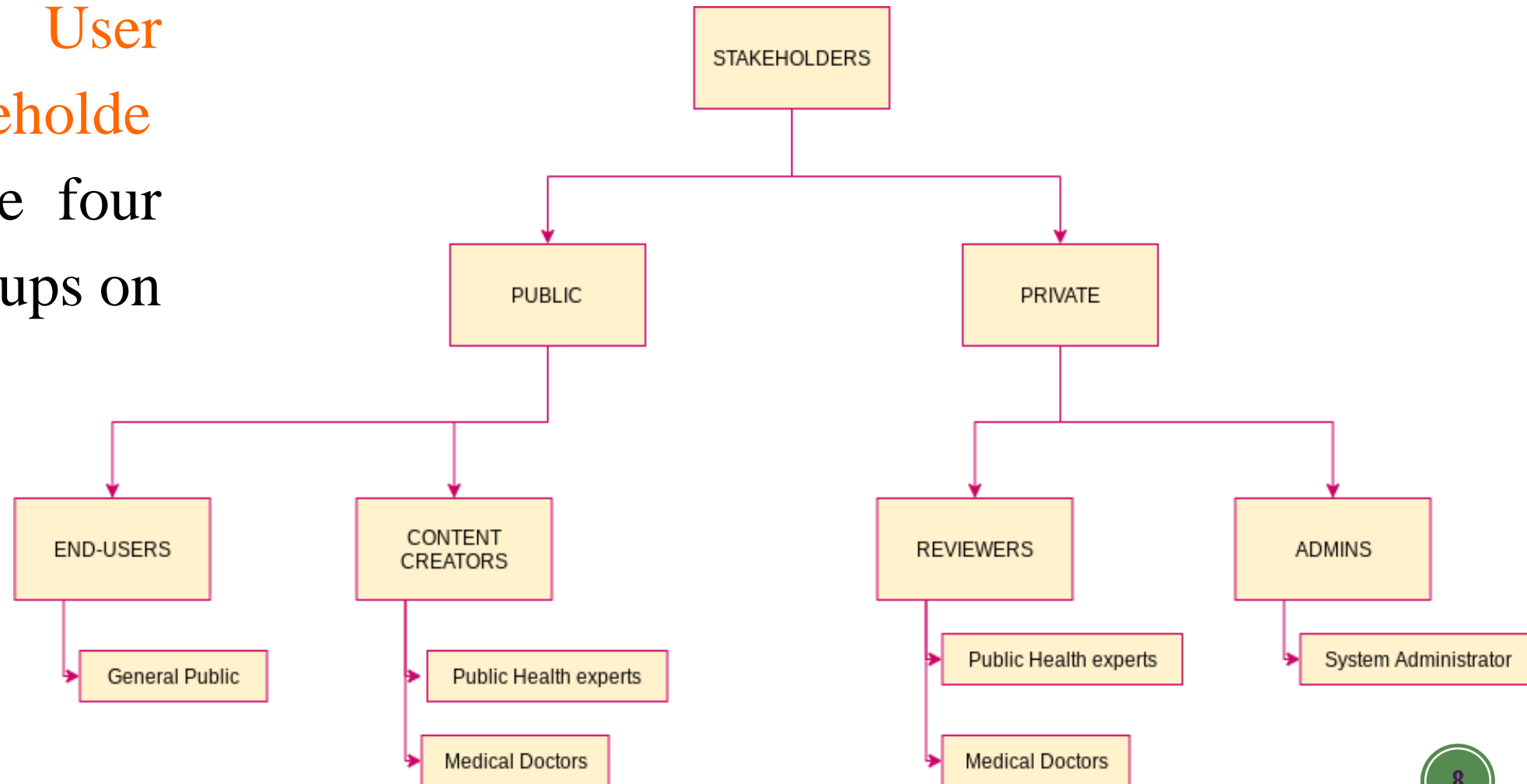
4.1 OBJECTIVE 1 – PUBLIC HEALTH INFORMATION SYSTEM (PHIS)

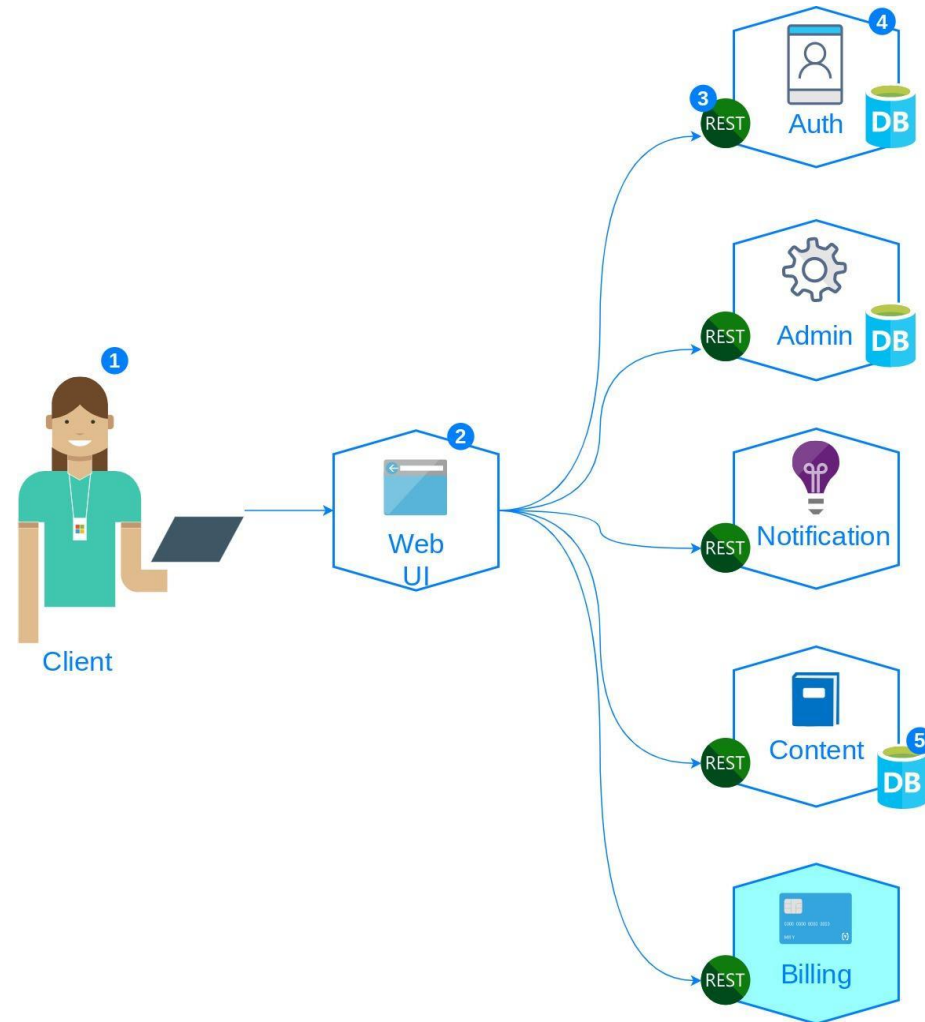▪ **PHIS User Groups/Stakeholders:** There are four key user groups on PHIS.

PHIS Stakeholders or User Roles

- The **PHIS Microservice Architecture** with services such as:
  - *Auth*
 - *Admin*
  - *Notification*
  - *Content*
  - *Billing*

1. **Client**: The end-user of the product

2. **The product**: FEDGEN Public Health Information System

3. **API**: REST over HTTP kind of API used in this deployment

4. **Microservices:** services running on separate individual VMs or containers.

5. **Database:** various selection of database, MySQL, MongoDB with different services sharing a single database.

PHIS Microservice Based Architecture

i) Authentication Service Use Case Diagram

Authentication Service Use Case
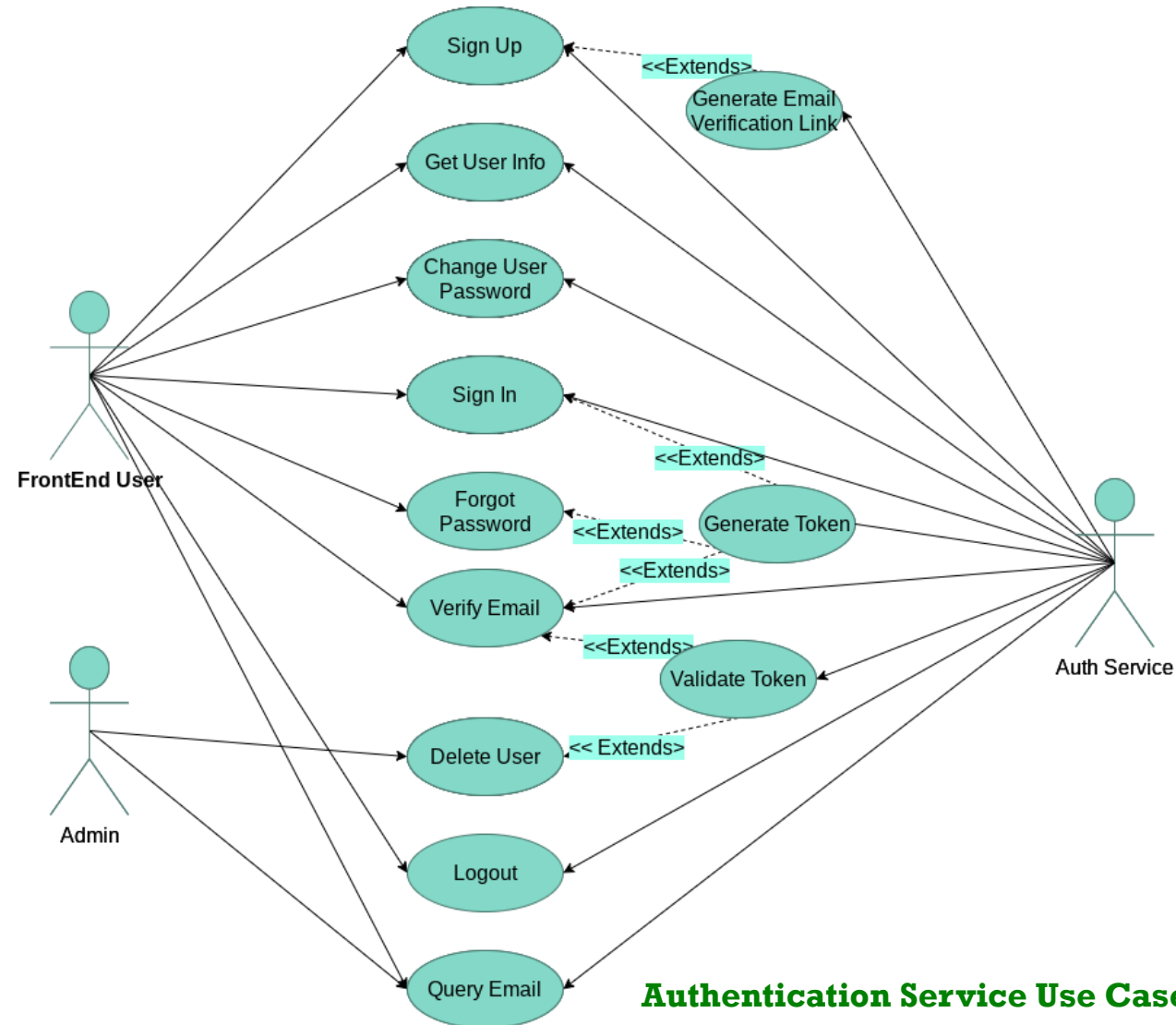
4.0 PRELIMINARY RESULTS CONT'D...

iii) Web User Interface (Frontend): Test deployment of the frontend at
https://phis.covenantuniversity.edu.ng

PHIS Frontend – Test Hosting on a University Sub-domain

4.2   OBJECTIVE 2 – EDGE AI SERVICE

Identification of Malaria Subtypes Biomarkers Using Deep Cascade Forest

Presenter: Dr. Itunuoluwa Isewon

4.2 IDENTIFICATION OF MALARIA SUBTYPES BIOMARKERS USING DEEP CASCADE FOREST

4.2.1 Aim & Objectives

Aim

- to identify malaria subtypes biomarker using the deep cascade forest model from whole blood transcriptomic gene expression dataset.

Objectives

- Data pre-processing and identification of differentially expressed genes (DEGs) to be used as labels for the machine learning.

- Train, test and evaluate the deep forest model and then compare with other ML (neural network, random forest, K-nearest neighbours, gradient boosting, decision trees and extra-trees classifiers) algorithms.

- Identification of biomarkers from the best performing model and further biological interpretations of the genes predicted.

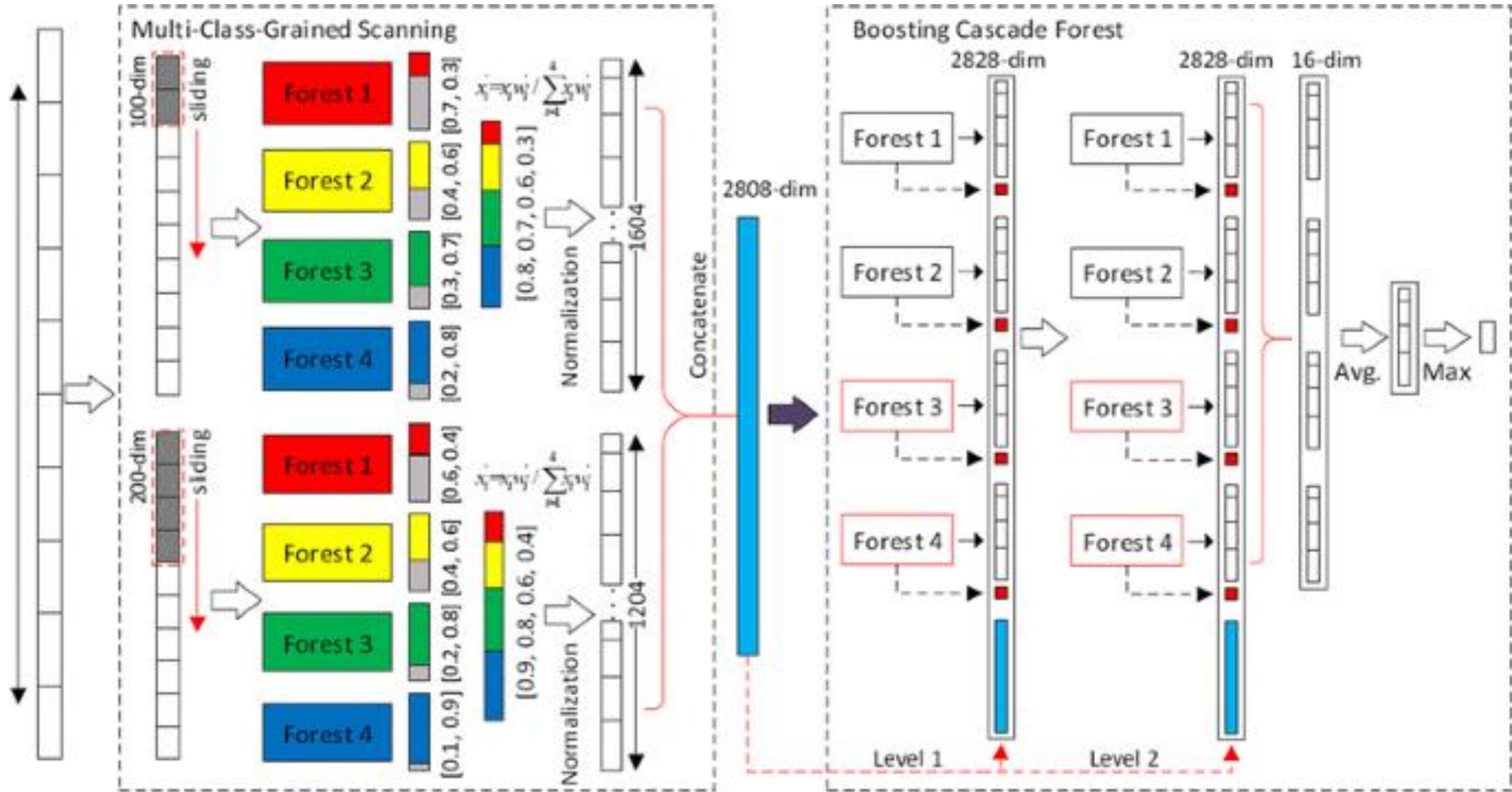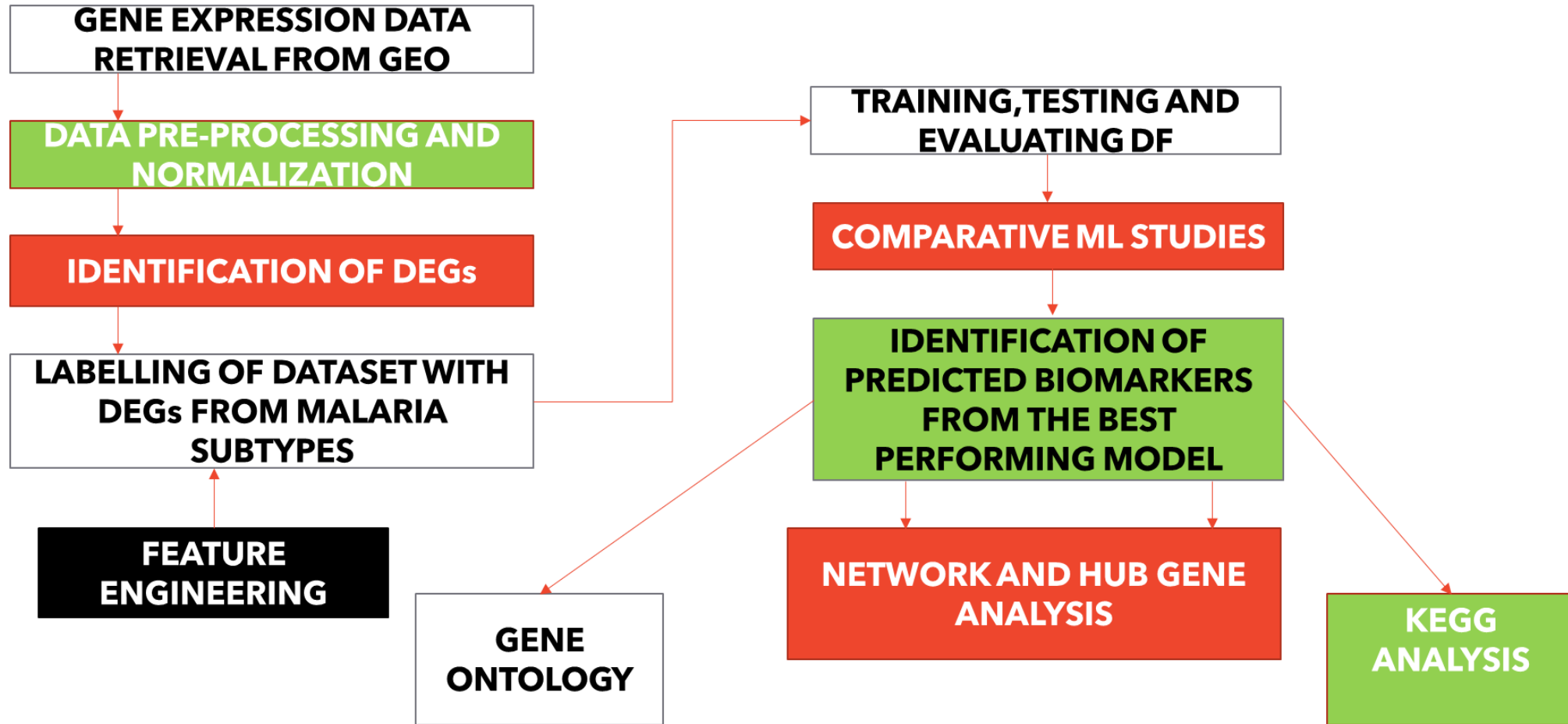# 4.2.2 Methodology for 2ⁿᵈ Objective

DECISION TREE

RANDOM FOREST

DEEP NEURAL NETWORK
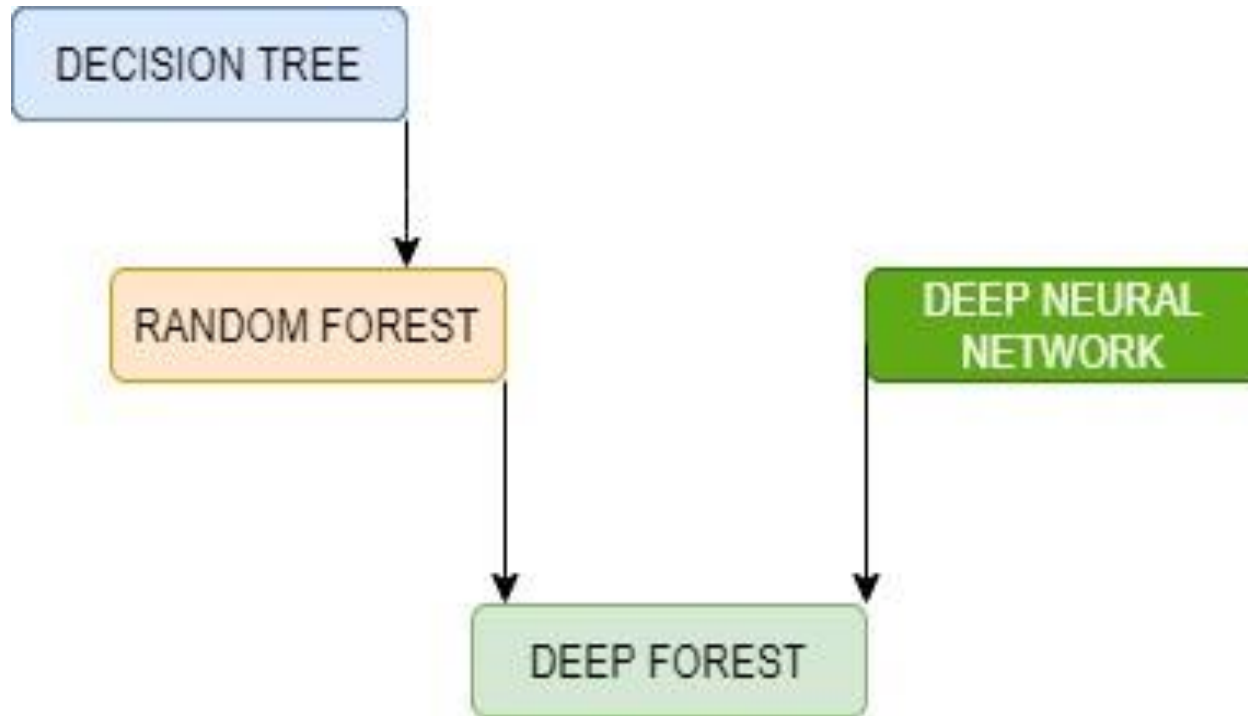
DEEP FOREST

Deep forest is a deep model that can :

- **Be trained without backpropagation**
- **Can learn from non-differentiable modules**

4.2.3 Workflow

GENE EXPRESSION DATA RETRIEVAL FROM GEO

DATA PRE-PROCESSING AND NORMALIZATION

IDENTIFICATION OF DEGs

LABELLING OF DATASET WITH DEGs FROM MALARIA SUBTYPES

FEATURE ENGINEERING

TRAINING, TESTING AND EVALUATING DF

COMPARATIVE ML STUDIES

IDENTIFICATION OF PREDICTED BIOMARKERS FROM THE BEST PERFORMING MODEL

GENE ONTOLOGY

NETWORK AND HUB GENE ANALYSIS

KEGG ANALYSIS

- The identified DEGs for each malaria subtype were used to label the data for a multi-class ML.
- The expression values were averaged and used as a single feature and more features were generated using various database and computational techniques such as protr, codonW, g:profiler, deeploc etc.
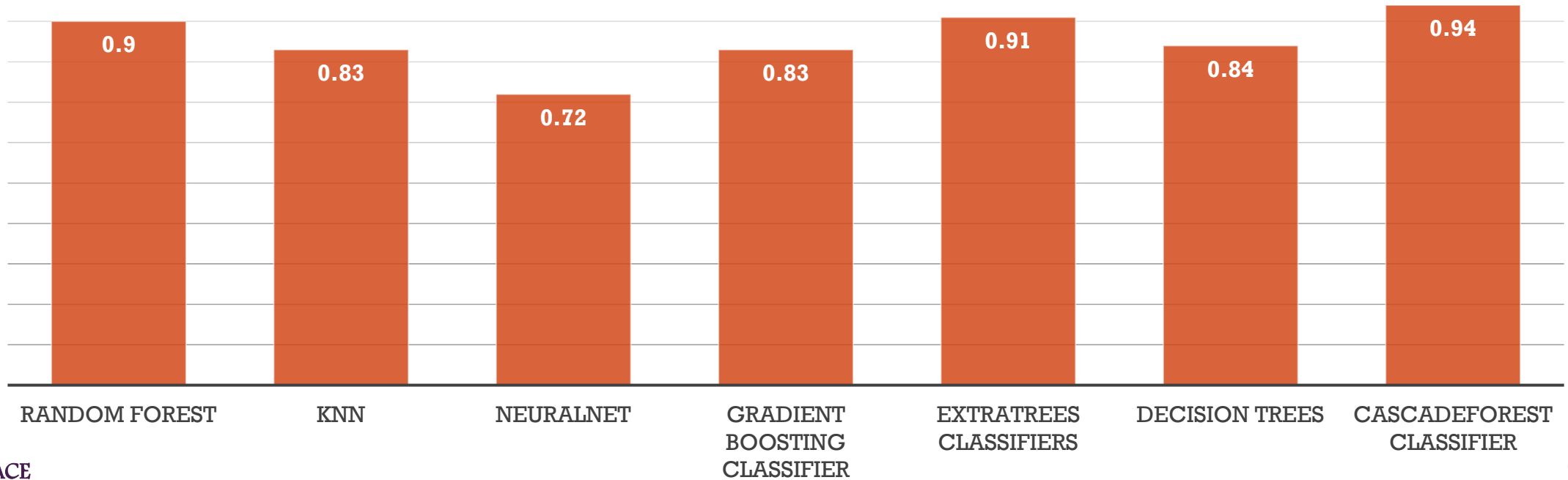
4

4.2.4 Preliminary Results

- A data split ratio of 80/20 was used for the dataset.
- An accuracy of 94 % was recorded for the deep cascade forest.
- A ROC-AUC score of 93% was also recorded for the Deep cascade forest.

Accuracy

0.9	0.83	0.72	0.83	0.91	0.84	0.94
RANDOM FOREST	KNN	NEURALNET	GRADIENT BOOSTING CLASSIFIER	EXTRATREES CLASSIFIERS	DECISION TREES	CASCADEFOREST CLASSIFIER

4.2.4 Preliminary Results Cont'd...
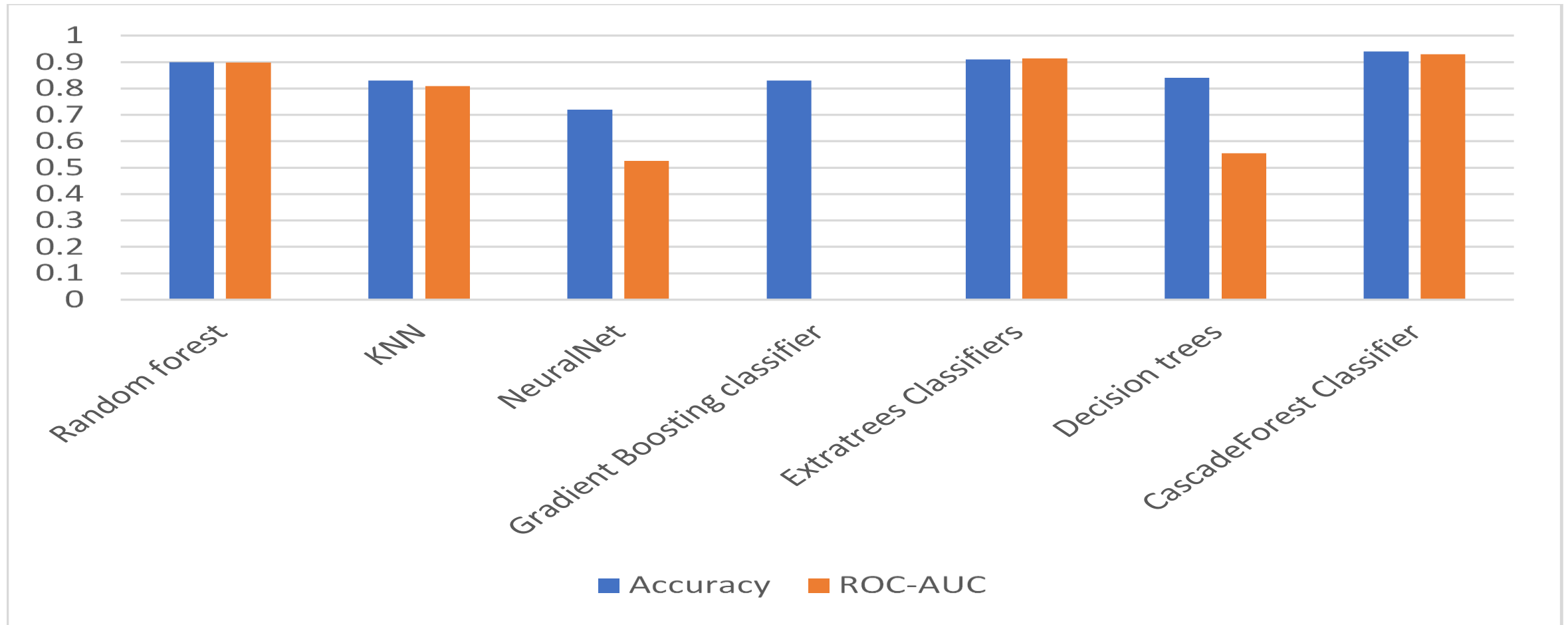
4.2.4 Preliminary Results Cont'd...

Subtypes	Asymptomatic(AS,1)	Cerebral(CR,2)	Uncomplicated(UC,3)	Anaemic(AN,4)
Number of DEGs	275	1779	706	372

S/N	Gene Symbol
1.	AK1
2.	RPL39
3.	PFDN5
4.	ECE1
5	SOCS3
6.	CTCF
8.	TERF2
9.	CCNG2

S/N	GENE SYMBOL
1.	DAP3
2.	MMP9
3.	CA4
4.	IL1R2
5.	HP
6.	SOCS3
7.	S100A12
8.	HP
9.	VNN1
10.	ARG1

S/N	GENE SYMBOL
1.	C1QB
2.	BCL2A1
3.	HP
4.	VCAN
5.	VNN1
6.	S100A12
7.	SOCS3
8.	ZFP36L2
9.	LY75
10.	RPS6KA5

S/N	GENE SYMBOL
1.	VCAN
2.	DEFA3
3.	IL1R2
4.	HP
5.	LTF
6.	MMP9
7.	HP
8.	S100A12
9.	VNN1
10.	ARG1

5.0 NEXT STEPS

- Completion and User Acceptance Evaluation of PHIS (Objective 1) – CApIC-ACE and CEA-MITIC.

- Cloud and Mobile App Implementation for Edge AI Services (Objective 2) – CEA-MITIC and CApIC-ACE .

- Requirements Elicitation, Design and Implementation of Research DataHub (Objective 3) - CApIC-ACE.

- Public Awareness Campaign for FEDGEN_HealthEduApp. - CApIC-ACE and CEA-MITIC.

- Partnerships within the DSTN, other ACE Networks and Sectoral Partners, for Upgrade and Uptake.

Thanks for your attention !

https://ace.covenantuniversity.edu.ng/
ace@covenantuniversity.edu.ng
+234 806 307 3579

fedgen@covenantuniversity.edu.ng
emmanuel.adetiba@covenantuniversity.edu.ng

@capicace
@CApIC ACE
@Capic Ace