

VINGROUP



<https://vinbigdata.org/>

# VINGROUP BIG DATA INSTITUTE

“

WHO IS BETTER  
PLACED TO ADDRESS  
THE CHALLENGES  
OF VIETNAM THAN  
OURSELVES?

---

**Professor Vu Ha Van**

SCIENTIFIC DIRECTOR

VINGROUP BIG DATA INSTITUTE



---

## 1. OVERVIEW

About Vingroup	06
About VinBigdata	08

---

## 2. SCIENTIFIC RESEARCH

Biomedical Informatics	14
Medical Imaging Analysis	16
Speech & Language Processing	18
Computer Vision	20

---

## 3. PRODUCT ECOSYSTEM

VinGen	24
VinDr	26
VinBase	28

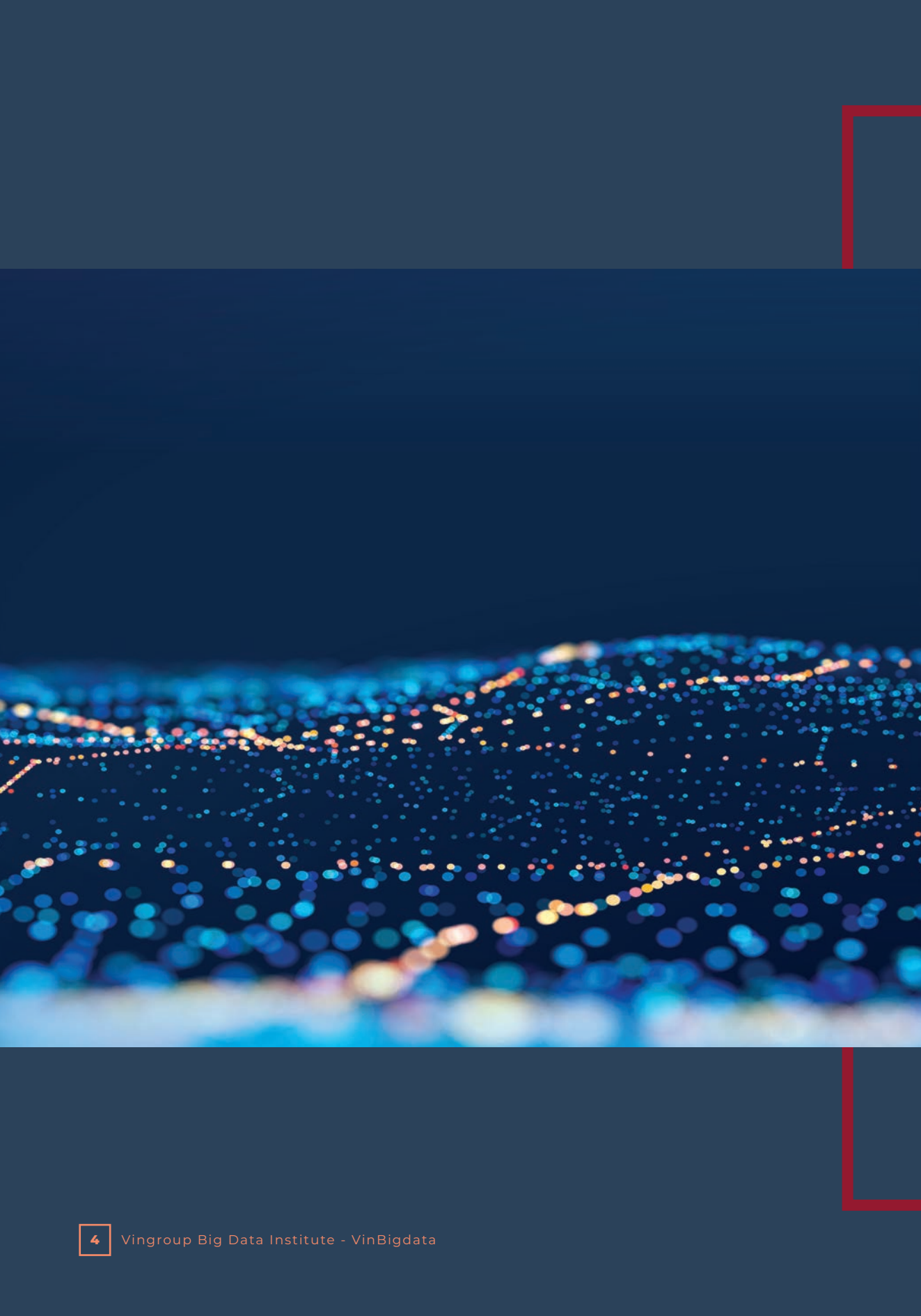
---

## 4. SCIENCE - TECHNOLOGY SPONSORING & TRAINING

Vingroup Innovation Foundation	36
Vingroup Human Resources Development Program in AI & Data Science	40

---

## 5. PARTNERS 42



# OVERVIEW

# 01



# "To create a better life for the Vietnamese people"

Since its establishment in 1993, Vingroup Joint Stock Company (Vingroup JSC) has developed to become the largest private enterprise in Vietnam and one of the biggest private conglomerates in Asia.

Vingroup follows its pioneering spirit to achieve market leadership and sustainable development, in order to become one of the most respected companies not only in Vietnam but also in the region. Vingroup will achieve this goal by improving product quality and service standards, by improving the quality of life for the Vietnamese people, and by raising the reputation of the nation in the world.



## TOP 10

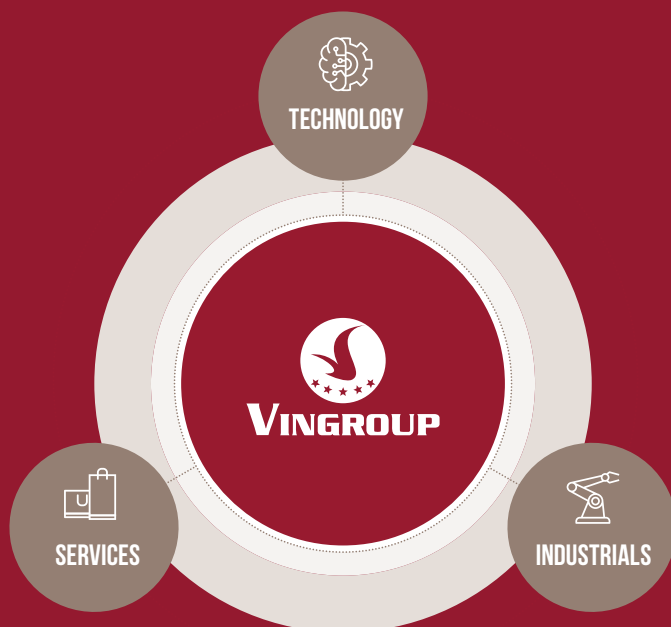
Largest Companies in  
Vietnam 2018 - VNR 500



Business size

## ~50.000

EMPLOYEES







*Landmark 81 - The highest building in Vietnam*

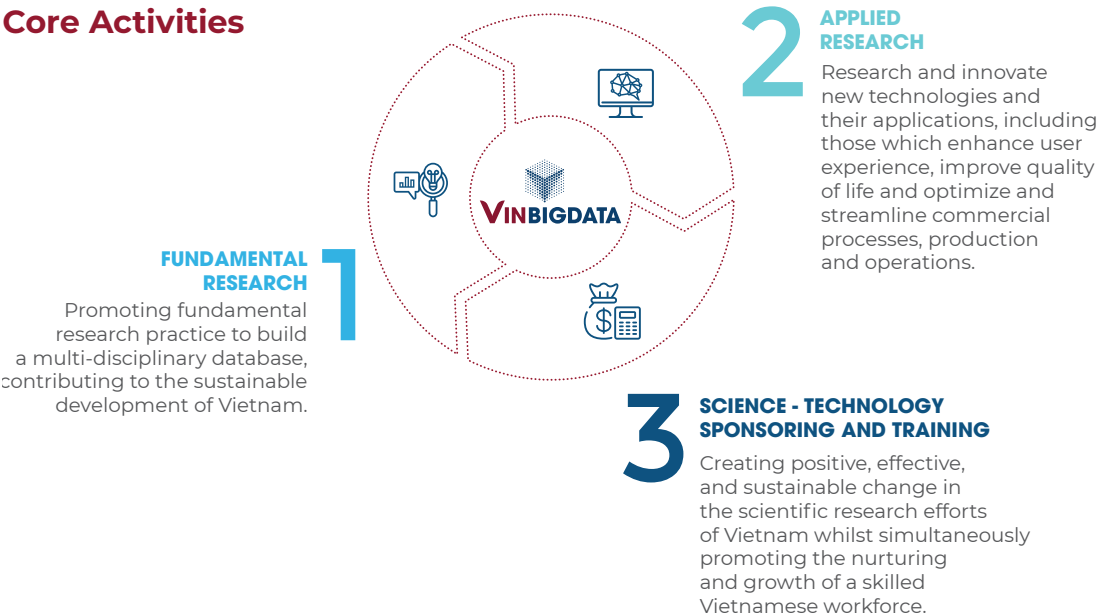


Vingroup Big Data Institute (VinBigdata) focuses specifically on the field of Data Science. The overarching vision of VinBigdata is to become the leading research institute in Vietnam, following and pioneering world-leading standards. VinBigdata was founded in August 2018 as part of a key strategic development for Vingroup's goal of becoming a world-class Technology - Industrials - Services corporation, with technology taking center stage.





## Core Activities



## Vision

**“ To become a world-class research institute in Technology and Data Science which leads the way in Vietnam**

## Mission

**Vietnamese Technology empowers Vietnam’s Future**



Bring global knowledge applied to the local context, VinBigdata aims to investigate, develop, and innovate novel products and solutions to strengthen the growing scientific and technological capacity of the nation. With the ultimate goal of building a solid foundation for the sustainable development of Vietnam's society and for the benefit of future generations, VinBigdata's scientific team specializes in research of fundamental and applied sciences.



# SCIENTIFIC RESEARCH

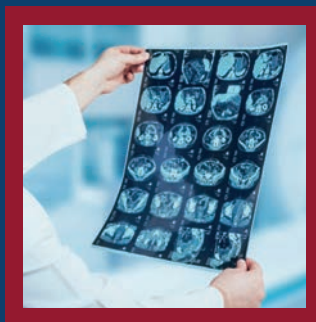
# 02



## BIOMEDICAL INFORMATICS

### Fundamental & Applied Research

1. Build biomedical databases for Vietnamese population
2. Develop advanced computational methods for biomedical data
3. Build systems for management and analysis of biomedical data
4. Develop solutions to predict disease risks and adverse drug reactions



## MEDICAL IMAGING ANALYSIS

### Fundamental & Applied Research

1. Chest X-ray interpretation
2. Mammography interpretation
3. Lung CT interpretation
4. Abdomen CT interpretation
5. Brain CT interpretation
6. Brain MRI interpretation
7. Bone X-ray interpretation

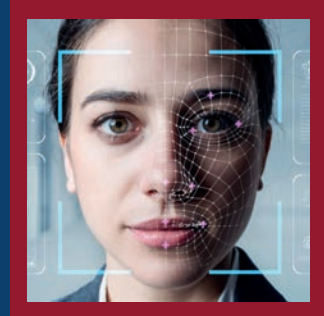




## **SPEECH & LANGUAGE PROCESSING**

### **Applied Research**

1. Speech Recognition
2. Speech Synthesis
3. Chatbot / Voicebot / Virtual Assistance
4. Sentiment Analysis
5. Machine Translation



## **COMPUTER VISION**

### **Applied Research**

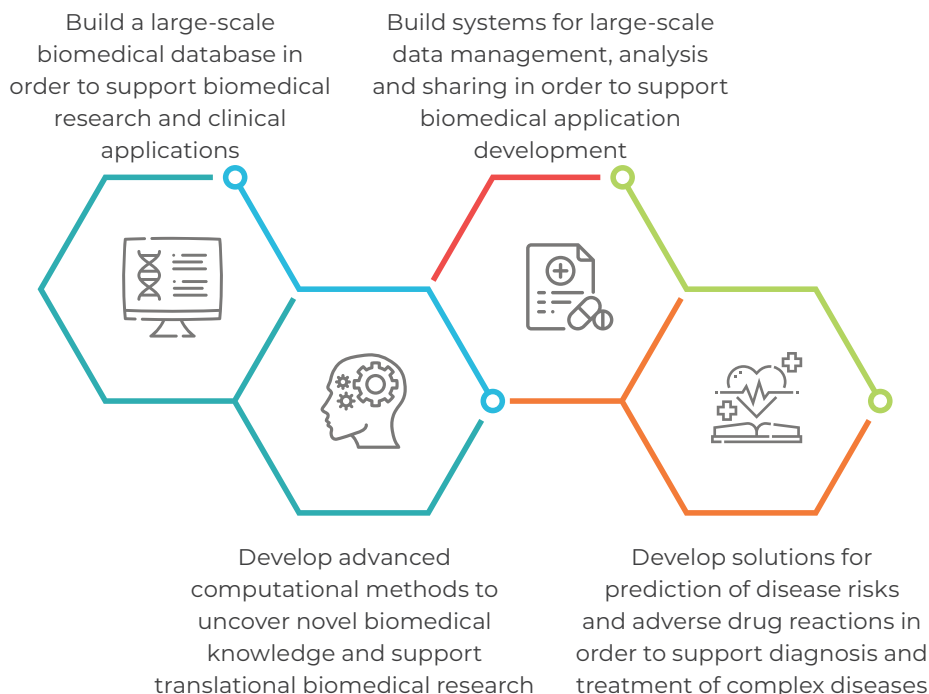
1. Face Recognition & Facial Attribute Analysis
2. Object Detection and Analysis
3. Optical Character Recognition (OCR)



# BIOMEDICAL INFORMATICS

Biomedical Informatics is one of the main research areas in VinBigdata, which aims at developing and applying computational methods and tools to analyze large-scale biomedical data to improve screening, diagnosis, and treatment of diseases.

## Perspectives on R&D



## A HUMAN GENOME PROJECT FOR 1000 VIETNAMESE PEOPLE

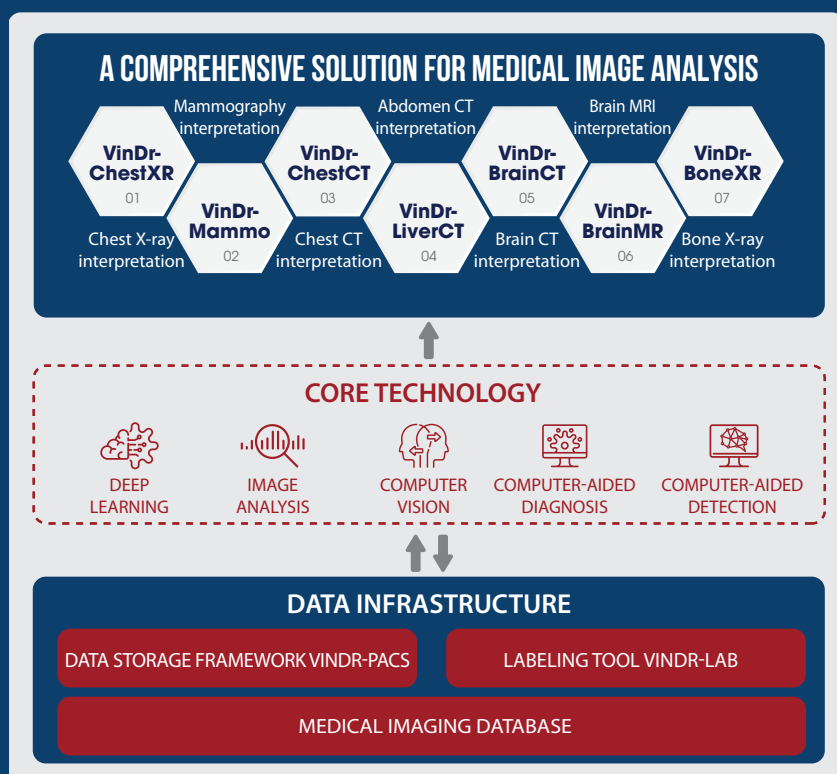
Vietnam ranks 15th among the most populous countries in the world. Despite the large population, current genetic research on Vietnamese population is often based on widely-used genetic databases which have very little information related to Vietnamese. Consequently, building a Vietnamese human genetic variation database is of paramount importance to develop our understanding of local genetic variations.

The project focuses on building the largest-scale database of Vietnamese human genetic variation to date. This will be accomplished through the sequencing of 1,000 Vietnamese people using Illumina's Next-Generation Sequencing, providing 30x depth of coverage. This data will then be analyzed to determine genetic variations, including single-nucleotide variants (SNVs), insertions and deletions (INDELs), and structural variants (SVs). This database is expected to establish a foundation for future research in genomics on Vietnamese population, including genetic disease risk prediction and pharmacogenomics; ultimately paving the way for the application of precision medicine in Vietnam.

# MEDICAL IMAGING ANALYSIS

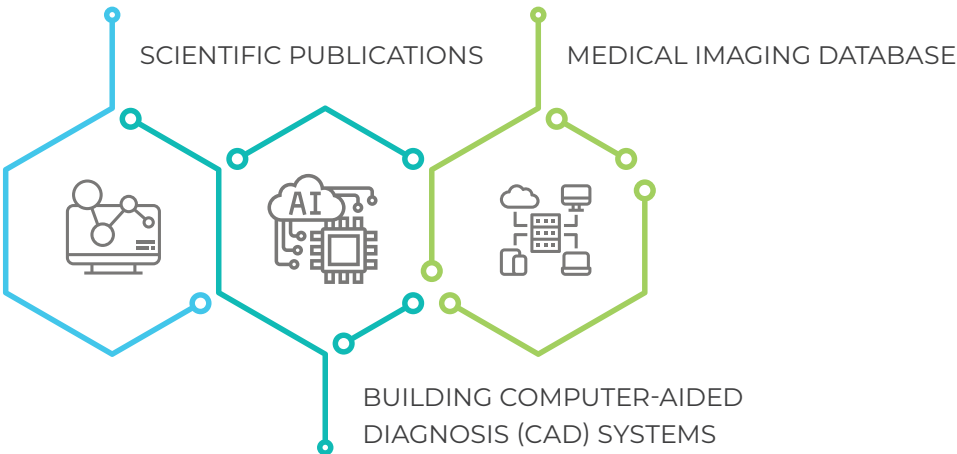
Medical Imaging Analysis combines both fundamental and applied research, which focuses on collecting, processing, analyzing, and understanding medical data. The main goal is to assist radiologists or clinicians in making fast and accurate diagnosis; therefore improving the patient care and facilitating effective clinical workflows in Vietnam.

By leveraging data-driven machine learning-based approaches, this research area focuses on processing and analyzing various medical imaging modalities such as X-ray, Mammography, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) as well as other types of medical data such as clinical or pathology information. These studies aim to screen and diagnose several popular and dangerous diseases in Vietnam, especially for cancers such as lung, breast and liver cancers, etc.





# Perspectives on R&D



# Research Achievements



**No. 1**

in CheXpert competition  
organized by Stanford  
University in 2019



**No. 1**

in Abnormal Image  
Detection in Endoscopy Videos  
(EndoCV), 2020



**TOP 3**

in Pulmonary Embolism  
Detection Challenge,  
organized by the Radiological  
Society of North America  
(RSNA), 2020



**TOP 10**

in Intracranial Hemorrhage  
Detection Challenge on CT scans,  
organized by the Radiological  
Society of North America  
(RSNA), 2019

# SPEECH & LANGUAGE PROCESSING

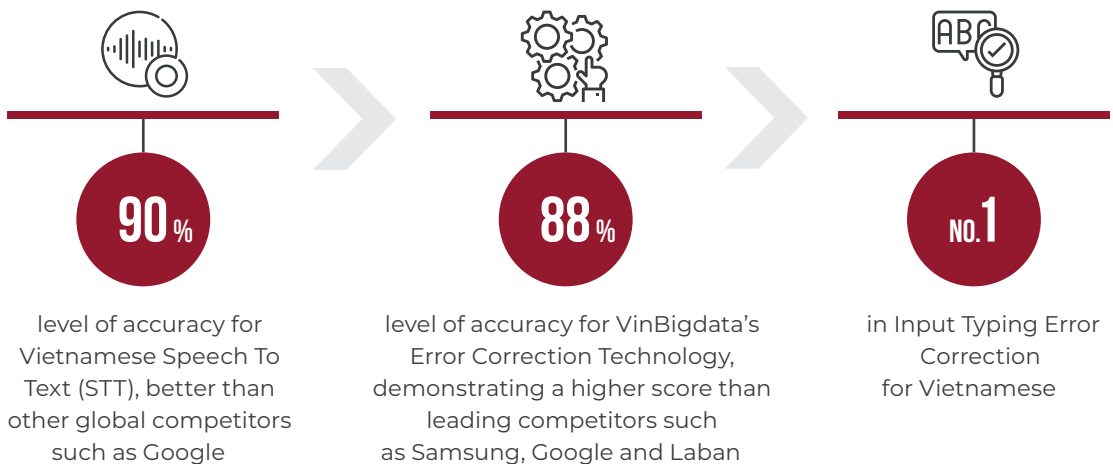
Speech and Language Processing is an area of applied research in artificial intelligence of VinBigdata, which aims to enable computer systems to listen, understand and express human language via text and/or speech.

At VinBigdata, the Speech and Language Processing department conducts primary research and development of technologies to process both structured and unstructured language data. VinBigdata's primary research orientations in this field include:

ENABLE COMPUTER SYSTEMS TO COMPREHEND HUMAN LANGUAGE

ENABLE COMPUTER SYSTEMS TO COMMUNICATE AND SUPPORT HUMANS IN SPECIFIC TASKS

Following these research orientations, VinBigdata focuses specifically on studying problems of parsing, semantic and sentiment analysis, as well as knowledge-based systems, dialogue systems, speech recognition and synthesis.



**01**

**Speech  
Recognition**

**02**

**Speech  
Synthesis**

**03**

**Chatbots/  
Voicebots/  
Virtual Assistance**

**04**

**Sentiment  
Analysis**

**05**

**Machine  
Translation**

# COMPUTER VISION

Computer Vision is a research area that focuses on collecting, processing, and analyzing data from images and videos, using exceptional data-driven approaches along with machine learning technology to effectively and practically address important societal issues.

The segment concentrates in the construction and creation of AI applications using image processing, computer vision, machine learning, and deep learning technologies. The outcome of such applications is to enhance the capabilities of computers in carrying out important recognition tasks, such as subject and action classification, detection, recognition, and tracing. Using the study results undertaken at the institute, VinBigdata is able to develop applications for customer data analysis, facial recognition, text recognition, text mining, vehicle classification, and many more.



93%

Applications demonstrating 93% accuracy in business card reader, extraction, classification, and directory storage of data. This outperforms similar products in the market



99%

Face recognition and analysis with 99% accuracy



97%

Up to 97% accuracy in character recognition for Vietnamese and English documents, which is currently the highest in the market



## Perspectives on R&D



### FACE RECOGNITION & FACIAL ATTRIBUTE ANALYSIS

- Face Detection
- Face Matching
- Face Anti-Spoofing
- Gender/Age Estimation
- Emotion Recognition
- Facial 3D Modelling
- Face Generator



### OBJECT DETECTION AND ANALYSIS

- Object detection
- Object recognition
- Multi-object tracking
- Re-Identification
- Gait recognition
- Action recognition
- Human attribute recognition



### OPTICAL CHARACTER RECOGNITION (OCR)

- Text detection/recognition
- Image segmentation
- Document layout analysis
- Character recognition
- Handwriting recognition
- Fraud detection
- Natural Language Processing







THROUGH THE RESONANCE  
OF DATA ALONG WITH VIETNAMESE  
HUMAN RESOURCES FROM ALL OVER  
THE WORLD, VINBIGDATA STRIVES TO  
CREATE PRODUCTS AND SOLUTIONS TO  
EFFECTIVELY SOLVE THE CHALLENGES  
FACING VIETNAMESE PEOPLE,  
LEADING US ON THE PATH  
OF SUSTAINABLE DEVELOPMENT.

---

**Dr. Dao Duc Minh**

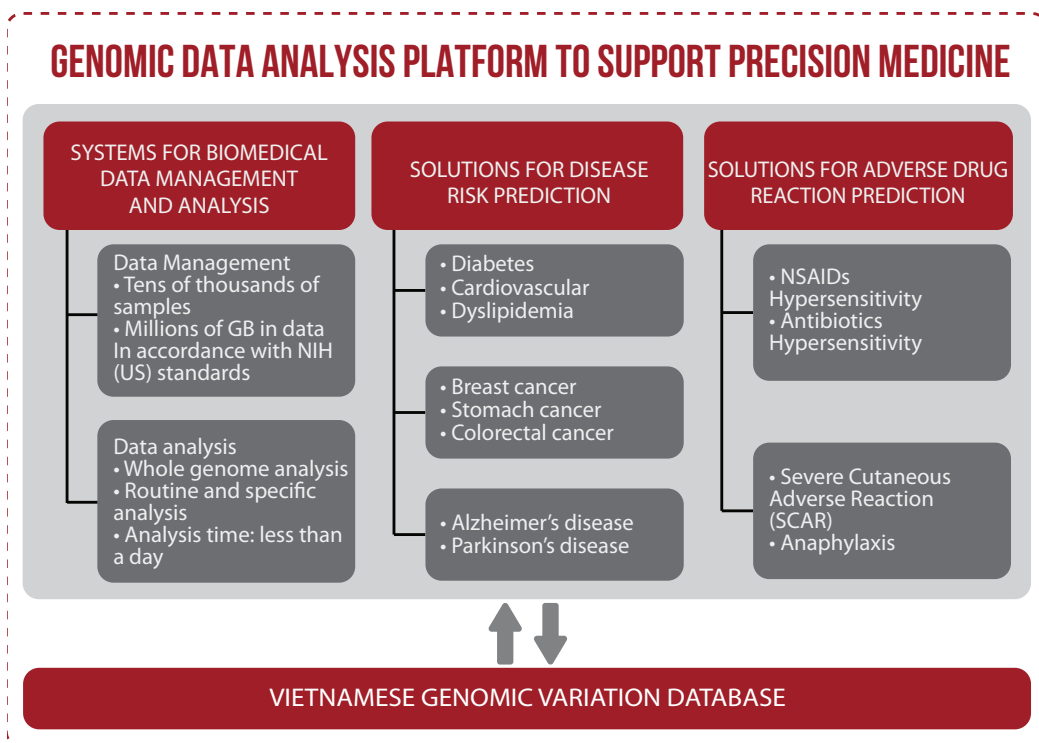
MANAGING DIRECTOR

VINGROUP BIG DATA INSTITUTE

# PRODUCT ECOSYSTEM 03

# VINGEN

## GENOMIC DATA ANALYSIS PLATFORM TO SUPPORT PRECISION MEDICINE



VinGen offers systems for management, analysis and sharing of genomic data, as well as solutions for disease risk prediction and adverse drug reaction prediction. These products are developed based on Vietnamese genomic data. As a result, this contributes towards the better diagnosis, treatment, and ultimately prognosis for a range of ailments, leading to better patient outcomes.

# 01

## Systems for Biomedical Data Management and Analysis

## 02 Solutions for Disease Risk Prediction

Development of solutions to predict the susceptibility to common diseases in Vietnam such as diabetes, cardiovascular disease, and cancer. These include the development of SNP chips to support doctors and genetic specialists in the screening, diagnosis and treatment of diseases.

These solutions have been developed based on 1,000 Vietnamese genomes data, and along with genotyping data of 3,000 people with genetic diseases undertaken by researchers in VinBigdata. The outcomes are expected to produce results with a higher level of precision than other available commercial products generally designed for many populations around the world.

## 03 Solutions for Adverse Drug Reaction Prediction

Development of solutions for predicting adverse drug reactions such as NSAID hypersensitivity, antibiotic hypersensitivity, severe cutaneous adverse reaction (SCAR), and anaphylaxis, among others. These include the development of test kits to support doctors and genetic specialists in the prescription of medicines.

The solutions have been developed based on 1,000 Vietnamese genomes data, and along with genotyping data of 4,000 people with ADR conducted by researchers in VinBigdata. The outcomes are expected to produce results that better fit Vietnamese genomic characteristics than other globally commercial products of the like.

This is currently the largest system for biomedical data management, analysis and sharing in Vietnam, using most advanced technologies in the world. This system is expected to become one of the most valuable reference portals for applied biomedical research and development, benefiting the community of researchers and professionals both in Vietnam and around the world.

CAPABLE OF PROCESSING MILLIONS OF GIGABYTES OF DATA WITH TENS OF THOUSANDS OF SAMPLES

CAPABLE OF PROVIDING A WHOLE-GENOME-ANALYSIS SERVICE WITH HIGH ACCURACY IN LESS THAN A DAY

SUPPORT BIOMEDICAL RESEARCHERS AND DOCTORS AND GENETIC SPECIALISTS IN DETERMINING DISEASE RISKS AND DRUG SIDE-EFFECTS



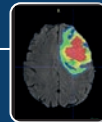
# VINDR

VinDr is a comprehensive solution for medical image analysis that integrates Artificial Intelligence (AI) into a Picture Archiving and Communication System (PACS) to assist radiologists in making fast and precise diagnoses. It aims to be an AI-assisted tool for medical doctors, which helps to improve the patient care and public health.

**VinDr-BrainCT**  
Brain CT  
interpretation



**VinDr-BrainMR**  
Brain MRI  
interpretation



**VinDr-ChestXR**  
Chest X-ray  
interpretation



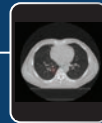
**VinDr-Mammo**  
Mammography  
interpretation



**VinDr-LiverCT**  
Abdomen CT  
interpretation



**VinDr-ChestCT**  
Chest CT  
interpretation



**VinDr-BoneXR**  
Bone X-ray  
interpretation



## VinDr-Mammo

85%

Mean  
accuracy  
reaches  
above 85%

<5s

The processing  
time is less  
than 5 second  
per study

13

Detect and  
localize 13  
common  
ammographic  
findings

## VinDr-ChestXR

90%

Mean  
accuracy  
reaches  
above 90%

<1s

The  
processing  
time is less  
than 1 second  
per study

28

Detect and  
localize 28  
common  
thoracic lesions  
and diseases

The VinDr-Mammo and VinDr-ChestXR have been deployed successfully for trial at three major hospitals in Hanoi, including 108 Military Central Hospital, Hanoi Medical University Hospital, and Vinmec Times City International Hospital. The solution was highly appreciated by medical doctors.

## PACS Features

MANAGE STUDIES IN DICOM FORMAT (X-RAY, CT, MRI)

BE EASILY INTEGRATED WITH SCANNERS OR PACS

DICOM VIEWER

SUPPORT RADIOLOGISTS TO MODIFY AND APPROVE DIAGNOSTIC RESULTS AS WELL AS GENERATE MEDICAL REPORTS

## AI Features

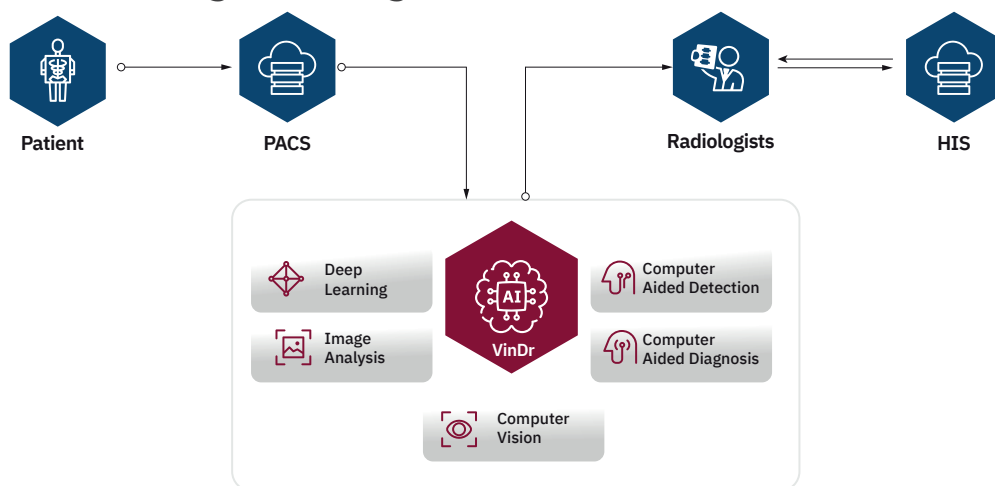
PROVIDE BOTH DISEASE LABELS AS WELL AS BOUNDING BOX ANNOTATIONS FOR LOCALIZING LESIONS

AUTOMATICALLY DIAGNOSE MULTIPLE STUDIES IN REAL-TIME

MODULES AVAILABLE: VINDR-CHESTXR AND VINDR-MAMMO

UPCOMING MODULES: VINDR-CHESTCT, VINDR-LIVERCT, VINDR-BRAINMR AND VINDR-BONEXR

## Outstanding technologies



## Outstanding advantages



Work independently or be two-way integrated with HIS/RIS/EMR/PACS/



Support multi-site diagnosis, perform analyzing multiple studies at the same time, across multiple devices



Flexible deployment, able to expand or scale down easily with cloud computing technology

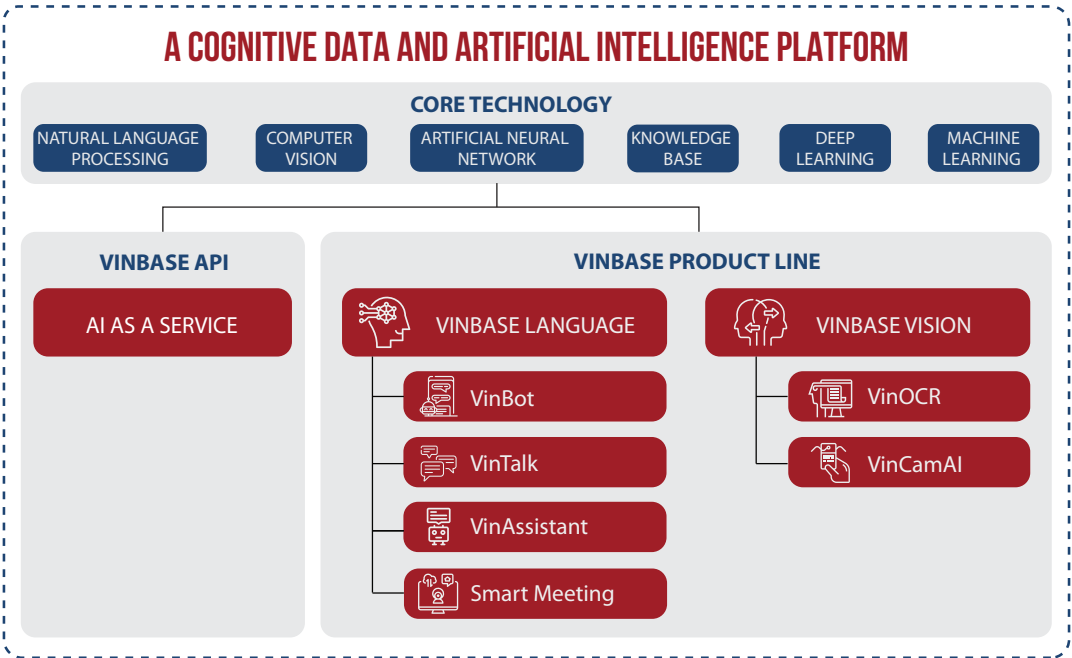


Meet the HL7 FHIR standard for information exchange between information technology systems according to the standard of the Ministry of Health

# VINBASE



VinBase is a cognitive data and artificial intelligence platform, which has been effectively developed based on several core technologies in natural language processing and computer vision, in combination with a multidisciplinary database. Existing solutions developed through VinBase aim for user experience enhancement, improving quality of life, along with optimizing and streamlining commercial business processes.



## VINBASE API

VinBase API allows to integrate AI-based core technologies into third-party applications, via cloud computing. API Services include: Speech Recognition, Speech Synthesis, Natural Language Processing, Machine Translation, Face Recognition, Optical Character Recognition, Handwriting Recognition, User Information Extraction and Authentication, and many more.

## VINBASE LANGUAGE

VinBase Language is a product line developed based on Natural Language Processing, providing chatbot, voicebot, virtual assistant and automated meeting notes to support business operation and optimization, as well as enhancing the experience of users. VinBase language products are multi-channel integrable and thus are well-suited to corporate governance systems. VinBase Language aims to provide an efficacious and user-friendly virtual assistant which is tailored not only to Vingroup's product ecosystem, but to the community in general.



### VinBot

VinBot provides a complete toolbox for chatbot building, configuration and customization for business purposes. It supports integration among social media channels (Facebook, Zalo...) and websites, in order to interact with end-users effectively. The key features of VinBot include:

- BUILD, MANAGE, AND SHARE CHATBOTS

- CONNECT TO MULTI COMMUNICATION CHANNELS

- PROVIDE 24/7 ONLINE SUPPORT

- PROVIDE OPERATIONAL STATISTICS AND REPORTING OF CHATBOTS



### VinTalk

VinTalk is a combination of Speech Processing (speech recognition, speech synthesis) and Natural Language Processing, thus creating a unique voicebot system that ensures accuracy of information recording and responds with speech automatically. The main features of VinTalk include:

- SPEECH RECOGNITION (SPEECH TO TEXT)

- COMPREHENSION OF USER INTENT INFERENCE

- SPEECH SYNTHESIS (TEXT-TO-SPEECH)





## Smart Meeting

Smart Meeting is an automated meeting minutes solution, using voice recognition and speaker recognition. The solution is integrated on mobile devices, smart TVs, smart speakers, and other forms of hardware. The unique features of Smart Meeting include:

MEETING VOICE RECORDING AND SPEECH-TO-TEXT

CONTENT RECOGNITION AND SEPARATION ACCORDING TO SPEAKERS



## VinAssistant in Healthcare

VinAssistant for doctors is a solution specifically designed for the healthcare sector, which looks to support physicians in undertaking image diagnosis, thus reducing the time required for administrative procedures like data inputting and reporting. VinAssistant has two main features as follows:

SUPPORT DIRECTLY INPUTTING DATA OF MEDICAL RECORDS AND REPORTS THROUGH SPEECH-TO-TEXT

SUPPORT MACHINE TRANSLATION BETWEEN ENGLISH AND VIETNAMESE OF MEDICAL RECORDS AND REPORTS EXTRACTED FROM MANAGEMENT SOFTWARE



## VINBASE VISION

VinBase Vision is a product line developed based on Computer Vision; offering identification and analysis, as well as extraction of images, objects, handwriting, documents, templates, and other forms of data. These applications are useful in areas including authentication and identity recognition, security surveillance, face recognition, vehicle counting, and automatic input. VinBase Vision solutions are easily integrated into existing systems of enterprises, thus leading to business operation optimization and improvement of user experience.



VinCamAI is a smart camera solution with the ability to detect and identify faces, analyze facial features (age, gender, emotions, and accessories), subject and vehicle recognition, behaviour recognition, and movement tracking.

### VinCamAI

#### CUSTOMER DATA

- Gender
- Age
- Foreign customers
- Behavior
- Emotion

#### MOVEMENT TRACKING

- Heat map
- Movement flow
- Crowded area
- Shopping time in specific sections

#### QUEUE MANAGEMENT

- Check-out time
- Check-out head count

#### HEAD COUNT

- Customer flow (day/hour)
- Total count of supermarket customers

#### AREA OF OPERATION

- Customer data in specific sections of supermarkets
- Restricted area warning



#### FACIAL RECOGNITION

- Face search
- VIP Identification
- Stranger warning

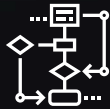


## VinOCR

VinOCR is a series of products supporting written information detection, identification and extraction from images. The solution can process formatted and non-formatted documents, tables and graphs, and handwriting; as well as easily customized, implemented and integrated into existing business systems.



Step 1:  
Identify the type  
of image to support



Step 2:  
Image Analysis

- Text detection
- Text reading
- Content analysis



Step 3:  
Convert image  
to text

### TYPE OF DOCUMENTS APPLICABLE

- Identity document
- Passport
- Driver license
- Vehicle Registration
- Household Registration Book
- Birth Certificate

96%

Accuracy

<1s

Process Speed

# SCIENCE TECHNOLOGY SPONSORING & TRAINING

# 04







Pioneer in sponsoring, training, and connecting the scientific and technological community, VinBigdata works towards changing mindset in doing research and promoting sustainable development of next generation scientists. We resonate the strength of Vietnamese knowledge and together create innovations to serve Vietnamese people.



# VINGROUP INNOVATION FOUNDATION - VINIF

Vingroup Innovation Foundation – VINIF was founded by Vingroup with the overriding mission of supporting organizations and individuals in scientific, technological and innovative research in order to bring about positive and sustainable changes for Vietnam.

Support Vietnamese scientists to engage in world-class research and technological solutions that benefit our nation

## **SCIENTIFIC AND TECHNOLOGICAL RESEARCH PROJECTS SUPPORT PROGRAM**



Enhance the quality of young and dynamic intellectuals by developing training programs, learning conditions and research project of international stature

## **MASTER TRAINING COLLABORATIVE PROGRAM**



## **POSTGRADUATE SCHOLARSHIPS SPONSOR PROGRAM**

Facilitate students, research fellows and doctors to strengthen their capabilities in study and innovative research, realizing their career paths in the scientific environment



## **SCIENTIFIC AND TECHNOLOGICAL CONFERENCES AND SEMINARS SPONSOR PROGRAM**

Promote collaborative exchanges among domestic and international researchers and enterprises



Methodical investment to create world-class scientific works and practically technological products for the nation is what VINIF is striving for. Each of our programs aim towards building an effective, professional research culture which allows researchers and young talents to devote themselves wholeheartedly to the pursuit of knowledge and creativity. With this, we create a sense of unity between the scientific, technological communities and enterprises of Vietnam.

---

**PhD. Phan Thi Ha Duong**

MANAGING DIRECTOR

VINGROUP INNOVATION FOUNDATION - VINIF

**VINGROUP INNOVATION FOUNDATION  
- VINIF**

9<sup>TH</sup> FLOOR, CENTURY TOWER,  
TIMES CITY, 458 MINH KHAI,  
HAI BA TRUNG, HA NOI

INFO@VINIF.ORG

[HTTPS://VINIF.ORG](https://vinif.org)



80+

Leading Professors  
and Doctors in VinIF's  
Scientific Council

100+

Universities and  
Research Institutes  
in our ecosystem

4.000+

Researchers and young  
intellectuals across the country  
in our network

30+

Scientific and Technological  
fields being invested and  
developed

3.000+

Scientists, postgraduates  
and research fellows benefit  
from our programs

4+

Key programs  
and activities







# VINGROUP HUMAN RESOURCES DEVELOPMENT PROGRAM IN **AI & DATA SCIENCE**







The Vingroup Human Resources Development Program in AI and Data Science was inaugurated in August 2020 with a view to build a team of high-quality technology engineers having the capacities to solve de facto problems and, in such a way, responding to increasing demand for technology human resources within the Corporation's entire ecosystem.

By focusing on AI and Data Science training, the program equips young engineers with application-oriented technological knowledge and offers opportunities to experience the real working environment via Vingroup's key technology projects. Trainee engineers are expected to become part of an elite workforce dedicated to exploring new technologies, promoting creativity and elevating science and technology of Vingroup and Vietnam in general.



### TRAINING PHASE



MACHINE  
LEARNING



COMPUTER  
VISION



NATURAL  
LANGUAGE  
PROCESSING



DEEP  
LEARNING



### PRACTICE PHASE

Experience the professional working environment and be part of world-class technology projects

# PARTNERS

## International Partners



# Domestic Partners

## Healthcare Sector



## Enterprise Sector





9<sup>th</sup> Floor, Century Tower, Times City, 458 Minh Khai, Hai Ba Trung, Hanoi