The Application of Machine Learning in Clustering Borderline Mild Cognitive Impairment among Aging Thai People Living with HIV

Mr. Akarin Hiransuthikul

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Health Research and Management
Department of Preventive and Social Medicine
FACULTY OF MEDICINE
Chulalongkorn University
Academic Year 2021

การประยุกต์ใช้ขบวนการการเรียนรู้ด้วยตนเองของคอมพิวเตอร์ในการจัดกลุ่มภาวะสูญเสียความสามารถของสมองในกลุ่มคนสูงอายุที่ติดเชื้อเอชไอวี

นายอัครินทร์ หิรัญสุทธิกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการวิจัยและการจัดการด้านสุขภาพ ภาควิชาเวชศาสตร์ป้องกันและสังคม
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564

Thesis Title	The Application of Machine Learning in Clustering Borderline Mild Cognitive Impairment among Aging Thai People Living with HIV
By	Mr. Akarin Hiransuthikul
Field of Study	Health Research and Management
Thesis Advisor	Professor VITOOL LOHSOONTHORN, Ph.D.
Thesis Co Advisor	THANAPOOM RATTANANUPONG, Ph.D.

Accepted by the FACULTY OF MEDICINE, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the FACULTY OF MEDICINE
(Associate Professor CHANCHAI SITTIPUNT)

THESIS COMMITTEE

..... Chairman
(Professor WIROJ JIAMJARASRANGSI, Ph.D.)

..... Thesis Advisor
(Professor VITOOL LOHSOONTHORN, Ph.D.)

..... Thesis Co-Advisor
(THANAPOOM RATTANANUPONG, Ph.D.)

..... External Examiner
(Nittaya Phanuphak Pungpapong, Ph.D.)

อัครินทร์ หิรัญสุทธิกุล : การประยุกต์ใช้ขบวนการการเรียนรู้ด้วยตนเองของคอมพิวเตอร์ในการจัดกลุ่มภาวะสูญเสียความสามารถของสมองในกลุ่มคนสูงอายุที่ติดเชื้อเอชไอวี. (The Application of Machine Learning in Clustering Borderline Mild Cognitive Impairment among Aging Thai People Living with HIV) อ.ที่ปรึกษาหลัก : ศ. ดร. นพ.วิฑูรย์ โล่ห์สุนทร, อ.ที่ปรึกษาร่วม : อ. ดร.ธนะภูมิ รัตนานุพงศ์

ผู้ที่อยู่ร่วมกับเอชไอวีจำนวนมากมีความบกพร่องของสมรรถนะทางสมอง รายละเอียดของการสูญเสียความสามารถของสมองนี้ยังมีไม่มากนัก ขบวนการการเรียนรู้ด้วยตนเองของคอมพิวเตอร์สามารถช่วยให้เห็นข้อมูลกลุ่มย่อยที่ซ่อนอยู่ การศึกษานี้มีวัตถุประสงค์เพื่อค้นหากลุ่มย่อยของคนไทยสูงอายุที่อยู่ร่วมกับเอชไอวีที่มีภาวะสูญเสียความสามารถของสมองเล็กน้อยโดยการใช้ขั้นตอนการเรียนรู้ด้วยตนเองของคอมพิวเตอร์ โครงการ HIV-NAT 207 รวบรวมผู้ที่อยู่ร่วมกับเอชไอวีสัญชาติไทยอายุ ≥50 ปี อาสาสมัครทั้งหมดได้เข้ารับการตรวจคัดกรองสมรรถนะทางสมองด้วย Montreal Cognitive Assessment (MoCA) การศึกษานี้ใช้ข้อมูลของอาสาสมัครที่ได้คะแนน MoCA ระหว่าง 23 ถึง 27 คะแนนสมรรถนะทางสมองแต่ละประเภทเป็นตัวแปรในการเข้าอัลกอริทึมการเรียนรู้ด้วยตนเองของคอมพิวเตอร์ K-means จากทั้งหมด 340 คน มีอาสาสมัครจำนวน 177 คน (52.1%) ที่ได้คะแนนระหว่าง 23 ถึง 27 โดยมีค่ามัธยฐานของอายุ 54 ปี (ค่าพิสัยระหว่างควอไทล์ 51-58) 118 คน (66.7%) เป็นเพศชาย ค่ามัธยฐานของ CD4 เท่ากับ 620 (ค่าพิสัยระหว่างควอไทล์ 489-795) เซลล์/ลูกบาศก์มิลลิลิตร และ 170 คน (96.1%) ตรวจไม่พบเชื้อไวรัสในเลือด จากการวิเคราะห์ด้วย K-means พบ 5 กลุ่มย่อยของอาสาสมัครทั้งหมดโดย 22.0% อยู่ในกลุ่มที่ 1 ความเสื่อมด้านความจำมากและภาษาเล็กน้อย 25.4% อยู่ในกลุ่มที่ 2 ความเสื่อมด้านการรับรู้มิติก่อสัมพันธ์/การบริหารจัดการ-ภาษา-ความจำเล็กน้อย 19.2% อยู่ในกลุ่มที่ 3 ความเสื่อมด้านการคิดเชิงนามธรรมปานกลางและการรับรู้มิติก่อสัมพันธ์/การบริหารจัดการ-ภาษา-ความจำเล็กน้อย 18.6% อยู่ในกลุ่มที่ 4 ความเสื่อมด้านภาษามากและความจำเล็กน้อย และ 14.7% อยู่ในกลุ่มที่ 5 ความเสื่อมด้านภาษาและการคิดเชิงนามธรรมมาก การได้มาซึ่งกลุ่มย่อยจะมีความสำคัญในการบ่งบอกความแตกต่างในทางคลินิกและการพยากรณ์โรคในอนาคตหรือไม่ เป็นประเด็นที่ควรมีการศึกษาติดตามระยะยาวต่อไปในอนาคต

สาขาวิชา	การวิจัยและการจัดการด้านสุขภาพ	ลายมือชื่อนิสิต .....
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

6470083530 : MAJOR HEALTH RESEARCH AND MANAGEMENT
KEYWORD: Aging, Cluster, Cognitive impairment, HIV, Machine learning, Thailand

Akarin Hiransuthikul : The Application of Machine Learning in Clustering Borderline Mild Cognitive Impairment among Aging Thai People Living with HIV. Advisor: Prof. VITOOL LOHSOONTHORN, Ph.D. Co-advisor: THANAPOOM RATTANANUPONG, Ph.D.

Many people living with HIV (PLWH) have cognitive impairment. Details of cognitive impairment subtypes are lacking. Unsupervised machine learning (ML) can reveal hidden subgroups within heterogeneous data. The study aimed to determine clusters of aging Thai PLWH with borderline cognitive impairment using unsupervised ML. HIV-NAT 207 study enrolled Thai PLWH aged ≥50 years. Cognitive performance was evaluated by the Thai-validated Montreal Cognitive Assessment (MoCA). This study included participants who scored between 23 and 27. The score of each cognitive domain served as cluster variables for the K-means algorithm. Among 340 PLWH, 177 (52.1%) scored between 23 and 27. Median age was 54 (IQR = 51-58) years, 118 (66.7%) were male, median CD4 was 620 (IQR = 489-795) cells/μL, and 170 (96.1%) were virally suppressed. K-means cluster demonstrated five clusters of all participants: 22.0% cluster 1 (marked memory with mild language impairment), 25.4% cluster 2 (mild visuospatial/executive function-language-memory impairment), 19.2% cluster 3 (moderate abstraction with mild visuospatial/executive function-language-memory impairment), 18.6% cluster 4 (marked language with mild memory impairment), 14.7% cluster 5 (marked language-abstraction impairment). A longitudinal study is warranted to identify differences in clinical significance and prognosis between each cluster.

Field of Study:	Health Research and Management	Student's Signature .....
Academic Year:	2021	Advisor's Signature .....
		Co-advisor's Signature .....

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors, Prof. Vitool Lohsoonthorn, M.D., Ph.D., and Thanapoom Rattananupong, Ph.D., who guided me throughout this project and advised me on all things during my transition period to be with the Department of Preventive and Social Medicine, Faculty of Medicine, Chulalongkorn University. I would also like to thank Ms. Pavita Dornkae for her kind assistance with the administrative process of the thesis preparation process.

I much appreciated the experience with Chaipat Chunharas, M.D., Ph.D., Mr. Waragon Phusuwan, Sedthapong Chunamchai, M.D., and their team at the Cognitive, Clinical & Computational Neuroscience, who introduced me to the world of machine learning and data science. It was great to work with them again on this project.

Anchalee Avihingsanon, M.D., Ph.D. has helped in several aspects of my career, including the collaboration to use the data from one of the best HIV research institutes in the world – The HIV Netherlands Australia Thailand Research Collaboration (HIV-NAT). I would also like to thank Ms. Tanakorn Apornpong, who helped me with the data extraction, and the HIV-NAT team for welcoming me with open arms.

Last but not least, I am forever grateful for the unconditional support my wonderful wife has given me for almost two decades since we have been together and for several warm hugs from our 2-year-old Milin-chan.

Akarin Hiransuthikul

TABLE OF CONTENTS

Page

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

CHAPTER 1

INTRODUCTION

Background and rationale

Thailand has approximately 11.6 million people aged 60 years or older, which accounted for 17.6% of the total population, almost a 50% increase from the previous decade (12.2%),[2] and double from two decades prior (9.6%) **(Figure 1)**.[3]

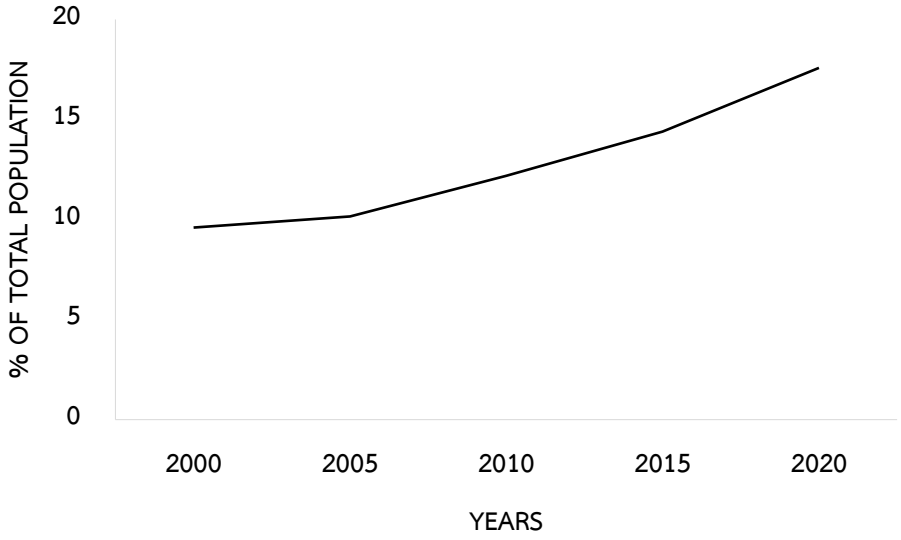


Figure 1. Percentage of total people aged 60 years or older each year.

Following this trend, Thailand is projected to enter a complete-aged society (a population which has the proportion of those ages 60 years or older exceeds 20%) by 2021 and a super-aged society (a population which has the proportion of those ages 60 years or older exceeds 28%) by 2031.[3] As the aging population increase, the number of age-related conditions such as dementia is expected to increase. This is of great concern as dementia is among the leading causes of death globally **(Table 1)**,[4]

a major cause of disability and dependency among the aging population, and would impact not only people with dementia but also their families. The ongoing coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2), has posed a unique impact on people with dementia as well. People with dementia are at high risk of SARS-CoV-2 acquisition because of the difficulty following the safeguarding procedure due to severe cognitive impairment. Living in care homes also facilitates viral transmission. Once infected, elderlies with dementia are more likely to experience severe symptoms than those without dementia. On the contrary, people living with dementia are at high risk of worsening behavioral and psychological symptoms of dementia (BPSD) due to social isolation during the pandemic.[5] Given that the COVID-19 pandemic is not expected to go away anytime soon, this poses new challenges for dementia care and urges for novel impactful management of these patients.

Identifying people at the stage of mild cognitive impairment (MCI), considered the pre-phase of dementia, is vital for early management or intervention. MCI refers to cognitive decline greater than expected for an individual's age and education level but does not interfere notably with daily life activities.[6] The prevalence of MCI increased substantially with age from an estimation of 6.7% among individuals aged 60-64 years to 25.2% among those older than 80 years. Up to 15% of individuals with MCI would develop dementia in 2 years. In addition, individuals with MCI were three times more likely to develop dementia than their age-matched counterparts.[7] Unfortunately, there are currently no pharmacological or dietary agents that show symptomatic cognitive benefit in MCI. The only recommendation for people diagnosed with MCI is regular exercise.[7] One of the reasons that may have led to previous negative results in clinical trials is the "heterogeneity" or multiple subtypes of MCI.

Table 1. Ten leading causes of mortality globally in 2019[4]

1	Ischemic heart disease
2	Stroke
3	Chronic obstructive pulmonary disease
4	Lower respiratory infections
5	Neonatal conditions
6	Trachea, bronchus, lung cancers
7	Alzheimer's disease and other dementias
8	Diarrheal diseases
9	Diabetes mellitus
10	Kidney diseases

People living with HIV (PLWH) process unique pathophysiology of cognitive impairment compared to HIV-negative counterparts. HIV enters the central nervous system (CNS) comportment prominently and quickly, as early as 8 days after infection.[8] Once inside the brain, HIV infects the residence cells of the CNS such as astrocyte, oligodendrocyte, and microglia, but not the neurons.[9] [10] Although neurons are not infected, they are still affected due to nearby inflammation. The exclusive infection of glial cells rather than neurons might be why subcortical domains are

predominantly impaired among PLWH,[11] which is different from the HIV-negative counterparts.

In summary, the number of PLWH with cognitive impairment is expected to increase as the number of aging population and life expectancy of PLWH increase. However, there is currently no effective intervention that can improve cognition or prevent or slow down the disease progression to dementia. The hidden "heterogeneity" or multiple subtypes of MCI may contribute to the negative results of previous clinical trials. Finally, PLWH have unique pathophysiology and anatomical localization that cause differed cognitive impairment compared to HIV-negative counterparts. The study to determine these hidden clusters of cognitive performance among PLWH will be the essential first step towards finding meaningful interventions management for PLWH with cognitive impairment.

Research questions

Primary research question*:* Can unsupervised ML reveal the heterogeneity (i.e. clusters) of aging Thai PLWH with borderline cognitive impairment?

Secondary research question: Are there any differences in demographic features between clusters of aging Thai PLWH with borderline cognitive impairment?

Objectives

Primary objective*:* To determine clusters of aging Thai PLWH with borderline cognitive impairment using unsupervised ML

Secondary objectives: To determine the differences in demographic features between clusters of aging Thai PLWH with borderline cognitive impairment

Research hypotheses

Hypothesis of primary objective: Unsupervised ML will be able to identify clusters of aging Thai PLWH with borderline cognitive impairment

Hypothesis of secondary objective: There will be some differences in demographic features, most likely education levels, between clusters of aging Thai PLWH with borderline cognitive impairment

Assumption

All data in this study are secondary data from Thai PLWH aged ≥50 years who were enrolled in the HIV-NAT 207 study and completed the MoCA.

Operational definition

PLWH are defined as those who had confirmed HIV infection by a licensed rapid HIV test or HIV enzyme or chemiluminescence immunoassay (E/CIA) test kit at any time prior to study entry and confirmed by a licensed Western blot or a second antibody test by a method other than the initial rapid HIV and/or E/CIA, or by HIV-1 antigen or plasma HIV-1 RNA viral load OR Documentation of HIV diagnosis in the medical record by a healthcare provider.

Cognition refers to all processes by which sensory input is transformed, reduced, elaborated, stored, recovered, and used.

Cognitive impairment refers to a decline in cognition greater expected from the person's age and education.

Machine learning (ML) refers to a subset of artificial intelligence that teaches computers to learn from experience, a natural process of humans and animals. ML used in this study was unsupervised learning.

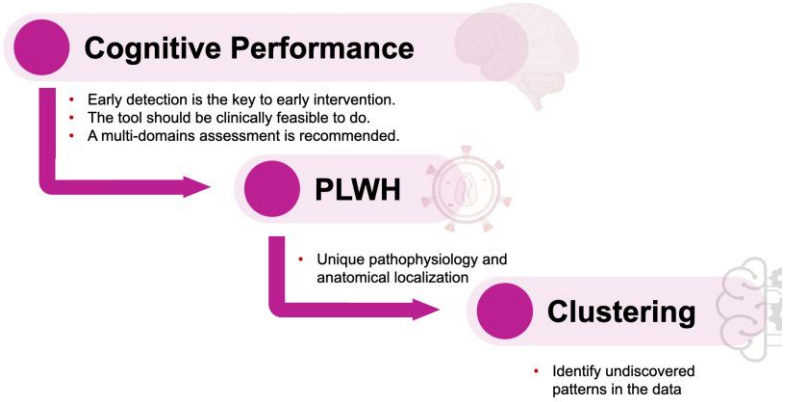
Expected benefit and application

Because all participants were enrolled in the prospective HIV-NAT 006 study, if the model can determine different clusters of cognitive performance among PLWH, those results can be integrated with the baseline characteristics from when the participants were initially enrolled in the HIV-NAT 006 study to determine associated factors with the development of the specific cluster; furthermore, they can be integrated with the current ongoing data to determine if any specific clusters are associated with future HIV and comorbidities outcome. These steps will be important evidence to support the integration of the cognitive assessment, which has been controversial in terms of the appropriate tools, using the MoCA to routine HIV care.

Obstacles and solution

No major obstacles were met. The data was extracted from the HIV-NAT database with ease, and the ML model was assisted by the expertise.

Conceptual framework



CHAPTER 2

REVIEW OF LITERATURE

Cognition and cognitive impairment

Cognition refers to all processes by which sensory input is transformed, reduced, elaborated, stored, recovered, and used.[12] Cognitive change as a normal process of aging has been well documented. As a person ages, the brain's cognitive abilities began to deteriorate; some are more resilient (e.g. vocabulary, reading, and verbal reasoning), some are more vulnerable (e.g. memory, attention, and processing speed).[13] However, in normal aging individuals, this often called age-related cognitive decline is subtle and does not significantly affect their abilities to perform daily tasks.

On the other hand, MCI and dementia indicate a decline in cognition greater than what would be expected from the person's age and education. MCI means that cognitive declines do not affect the person's ability to carry out everyday tasks (e.g., shopping, cooking, driving), while dementia indicates those cognitive difficulties are impacting the person's ability to complete everyday tasks. It is important to emphasize that both MCI and dementia are syndrome and do not imply the cause or the diagnosis. Alzheimer's disease is among the most common cause of MCI and dementia; other common causes include vascular MCI or dementia, frontotemporal lobe disease, and dementia with Lewy bodies (DLB). MCI can be thought of as a transitional state between normal aging and dementia.[14]

Based on the Key Symposium criteria since 2004, MCI can be categorized into four subtypes, based on two questions: is the memory domain involved, and is there more than one domain involved? The first question would categorize MCI into either amnestic or nonamnestic MCI, and the second question would categorize MCI into

either single or multiple domains. Thus, individuals with MCI can be classified as one of the four clinical subtypes: amnestic MCI-single domain, amnestic MCI-multiple domains, nonamnestic MCI-single domain, and nonamnestic MCI multiple domains **(Figure 2)**. Each type appeared to be associated with different etiologies and outcomes.[1] For instance, individuals with amnestic MCI were more likely to develop AD, while nonamnestic MCI were more likely to develop DLB.[15] The combination of clinical subtype and the presumed etiology could then be used to predict the type of dementia the individuals would most likely develop.[1 16] Because each subtype of MCI potentially processes diverse underlying pathology, without addressing this heterogeneity could potentially explain the reason why previous reports fail to demonstrate any impactful management or intervention for people with MCI. Although the clinical outcomes between those with and without multiple domain involvements are different, there has been limited information on what domains are involved in MCI-multiple domains.[17] In addition, it's more than likely that different phenotypes exist between those who score within the same range despite most cognitive assessments having binary outcomes.

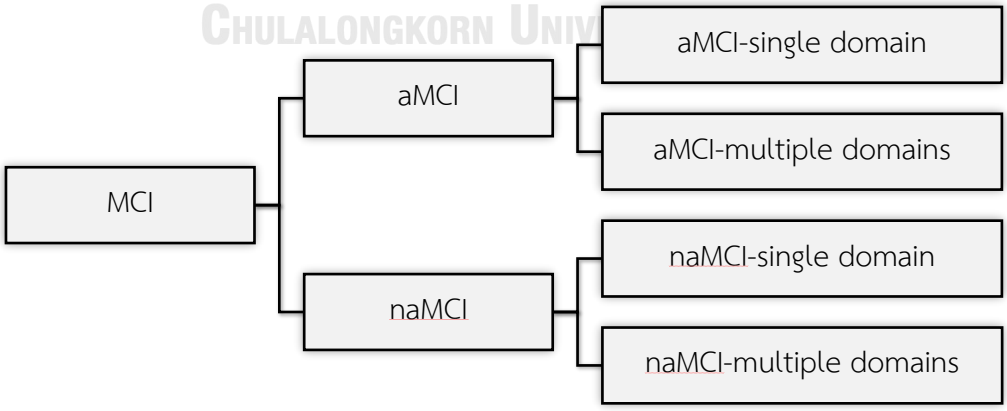


Figure 2. Based on the Key Symposium criteria,[1] mild cognitive impairment (MCI) can be categorized into 4 subtypes based on 2 aspects: memory domain involvement and number of domains involved.

Abbreviations: a, amnestic; MCI, mild cognitive impairment; na, nonamnestic

Montreal Cognitive Assessment

Multiple neuropsychological assessments are necessary to determine subtypes of cognitive impairment. Unfortunately, the complete evaluation takes a long time, in some cases up to several hours, and requires specialized healthcare personnel to conduct. Therefore, it is not feasible to perform such technical and time-consuming assessments in many settings, particularly resource-limited settings. In 2005, *Nasreddine et al* published a one-page 30-point cognitive screening tool that can be administered in 10 minutes to detect MCI called the Montreal Cognitive Assessment (MoCA).[18] The MoCA assesses multiple cognitive domains, including (1) visuospatial/executive function (clock-drawing task, three-dimensional cube copy, and Trail Making B task), (2) Naming (three-item confrontation naming task with low-familiarity animals), (3) attention (digits forward and backward, target detection using tapping, and serial subtraction task), (4) language (repetition of two syntactically complex sentences and phonemic fluency task), (5) abstraction (two-item verbal abstraction task), (6) delayed recall (short-term memory recall task), and (7) orientation (time and place). From the original study, the sensitivity and specificity of the MoCA in identifying MCI were 90% and 87%, respectively.[18] However, studies from different investigators have shown that the optimal cutoffs for the MoCA varied by race and ethnicities.[19] This also holds true to the Thai version of the MoCA **(Figure 3)**.[20] The Thai-validated version of MoCA showed sensitivity and specificity of 80% in MCI detection using a cutoff score of <25 by adding 1 point for subjects with ≤6 years of education. Therefore, the MoCA has many advantages, including time-saving, being widely available and accessible in many settings, and, most importantly, multiple domains assessment.

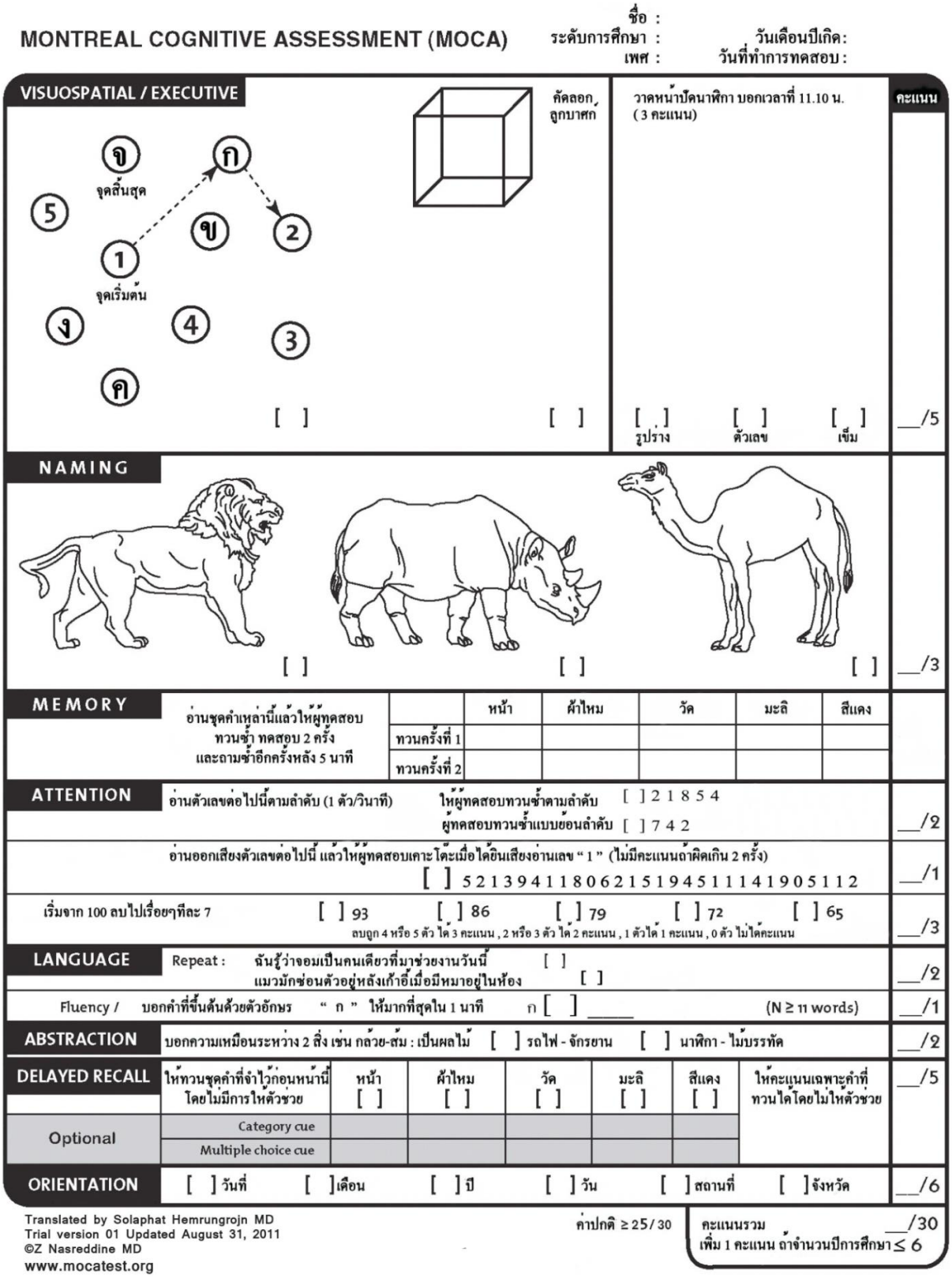


Figure 3. Thai-validated version of the Montreal Cognitive Assessment

Cognitive impairment and people living with HIV

One of the early AIDS-defining illnesses to be recognized is HIV-associated dementia (HAD). The incidence rate was 7% per year among individuals with advanced HIV.[21] After the invention of antiretroviral therapy (ART) and the application of combined ART, HAD is now rarely seen. In people living with HIV (PLWH) who have access to ART. Regardless, milder forms of cognitive impairment are being recognized among those who are virologically suppressed. In 2007, the terminology for these spectrums of cognitive impairment among PLWH was proposed as HIV-associated neurocognitive disorders (HAND), sometimes called the Frascati criteria.[22] HAND can be further categorized into 3 types based on their severity of objective neuropsychological findings and functional status: asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (MND), and HAD **(Table 2)**. In brief, ANI refers to PLWH with cognitive impairment but does not interfere with daily

Table 2. Classification of HAND

HAND type	Prevalence in combined ART-treated PLWH	Diagnostic criteria
ANI	30%	• Impairment in ≥2 neurocognitive domains (≥1SD) • Does not interfere with daily functioning
MND	20-30%	• Impairment in ≥2 neurocognitive domains (≥1SD) • Mild to moderate interference in daily functioning
HAD	2-8%	• Impairment in ≥2 neurocognitive domains (≥2SD) • Marked interference in daily functioning

Abbreviations: ANI, asymptomatic neurocognitive impairment; ART, antiretroviral therapy; HAD, HIV-associated dementia; HAND, HIV-associated neurocognitive disorders; MND, mild neurocognitive disorder; PLWH, people living with HIV; SD, standard deviation

function, MND refers to PLWH with cognitive impairment and mild-to-moderate interference of daily function and HAD refers to PLWH with marked cognitive impairment and marked interference of daily function. Interestingly, there were no differences in the prevalence of HAND between the pre-combined ART and combined ART era – range: 15-55% of PLWH.[11] However, the prevalence of milder forms of HAND has increased **(Figure 4)**. By linking to the traditional definition of MCI and dementia, ANI would be categorized as MCI (cognitive impairment without notable interference with daily life) and HAD would be categorized as dementia (cognitive impairment with notable interference with daily life). As for MND, this would depend on the degree of interference with daily life.

HIV can enter the central nervous system (CNS) comportment very prominently and at the early stage of infection. A study among Thai PLWH during acute HIV infection demonstrated that 83.3% have detectable CSF HIV RNA and can be detected early as 8 days after HIV acquisition.[8] HIV travels to the CNS by two main pathways: the Trojan-horse mechanism and directly through the breakdown of blood-brain barriers.[9] Once inside the brain, HIV will infect the residence cells of the CNS such as astrocyte, oligodendrocyte, and microglia, but not the neurons.[9 10] Although neurons are not infected, they are still affected due to nearby inflammation. The exclusive infection of glial cells rather than neurons might be the reason why subcortical domains are predominantly impaired among patients with HAND.[11] Therefore, screening batteries for HAND should focus on such domains – e.g. attention, fine movement, and executive function – and it is important to be aware of this unique anatomical localization compared to the general population.

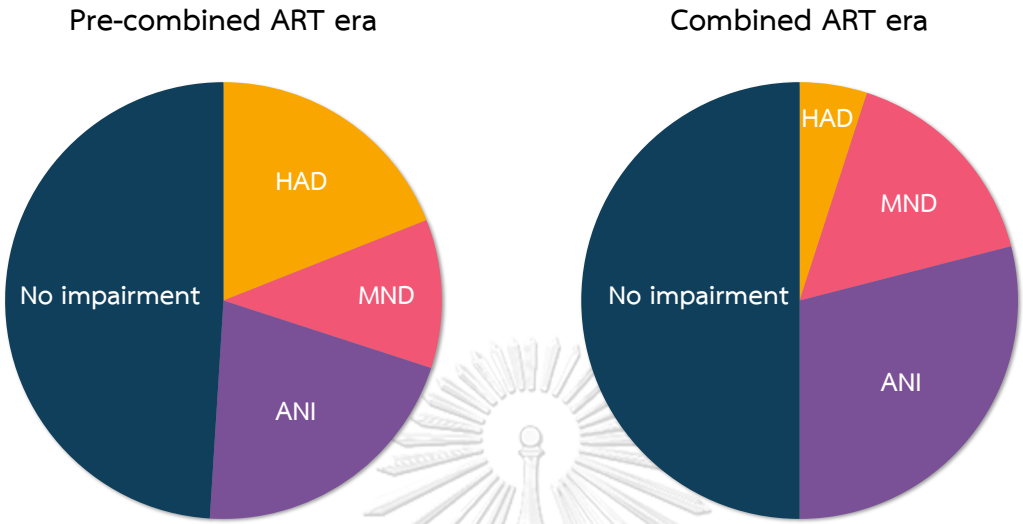


Figure 4. Prevalence of HAND between pre-combined ART and combined ART era. Since the introduction of combined ART in 1996, the proportion of HAND has remained unchanged, but the proportion of people with severe symptoms has declined.
Adapted from Saylor, D, et al. *Nat Rev Neurol* 2016;12(4):234-48.[14]
Abbreviations: ANI, asymptomatic neurocognitive impairment; ART, antiretroviral therapy; HAD, HIV-associated dementia; HAND, HIV-associated neurocognitive disorders; MND, mild neurocognitive disorder

Machine learning

Driven by the invention of the internet and an increase in computational power, Big Data was made possible and has become one of the most important assets of the 21st century.[23] Big Data is categorized by the 4 V's – volume (constant growing volume of data generated from different sources and which traditional databases cannot handle), variety (various platform of data collections), velocity (acquisition speed of the data and additionally the speed at which the data should be processed and analyzed), and value (the extraction of knowledge or patterns out of the raw data). However, it is very challenging, if not possible, to process this large

amount of data using the traditional approach. Therefore, machine learning (ML), a form of artificial intelligence (AI) most commonly used in healthcare settings, is introduced to tackle this data as it enables a robust interrogation of datasets to identify previously undiscovered patterns and relationships between different features in the data.[24]

ML teaches computers to learn from experience, a natural process of humans and animals. ML algorithms use computational methods to "learn" information directly from data without relying on a predetermined equation as a model. ML uses two types of techniques: supervised learning and unsupervised learning **(Figure 5)**. The concept of supervised learning is to build a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the answer to new data. In contrast, the concept of unsupervised learning is to finds hidden patterns or intrinsic structures in data.[25] In recent years, this data-driven approach has been increasingly integrated into the medical field. The application of ML algorithms to classify skin cancer and

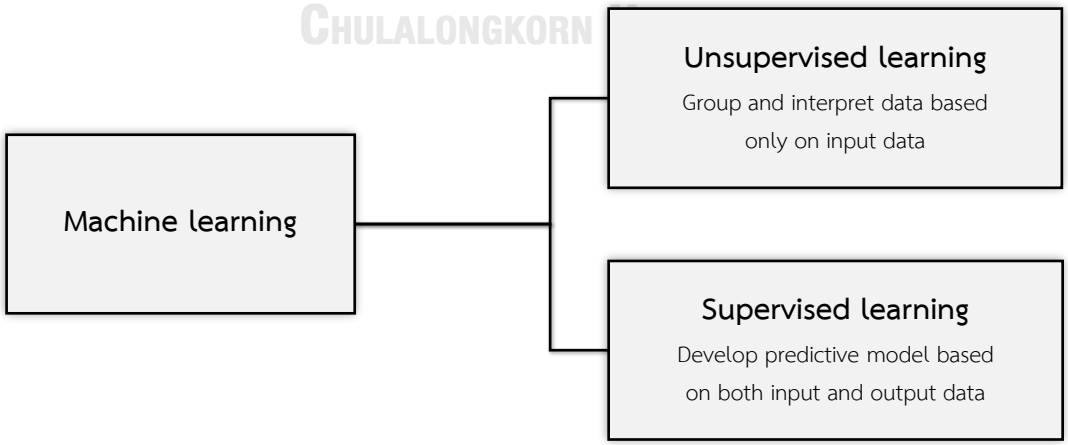


Figure 5. Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data.

predict the progression from pre-diabetes to type 2 are among the early success examples.[26 27] Unsupervised ML has been integrated into a handful of MCI studies, primarily to develop or evaluate the performance of diagnostic tools.[28-30] One study found that K-means clustering was able to reveal key patterns from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database using five sets of features – (1) Cognitive assessments and Ape genotype, (2) CSF, (3) MRI, (4) CSF and MRI, and (5) all of the above features – to aid early AD detection at the MCI stage.[31] However, applying this particular model is challenging, especially in resource-limited settings where access to advanced laboratory and imaging is limited. Among the most recent data presented at the 2021 American Academy of Neurology Annual Meeting, unsupervised ML demonstrated six clusters of aging Thai population with borderline MCI assessed by the MoCA **(Figure 6)**.[32]

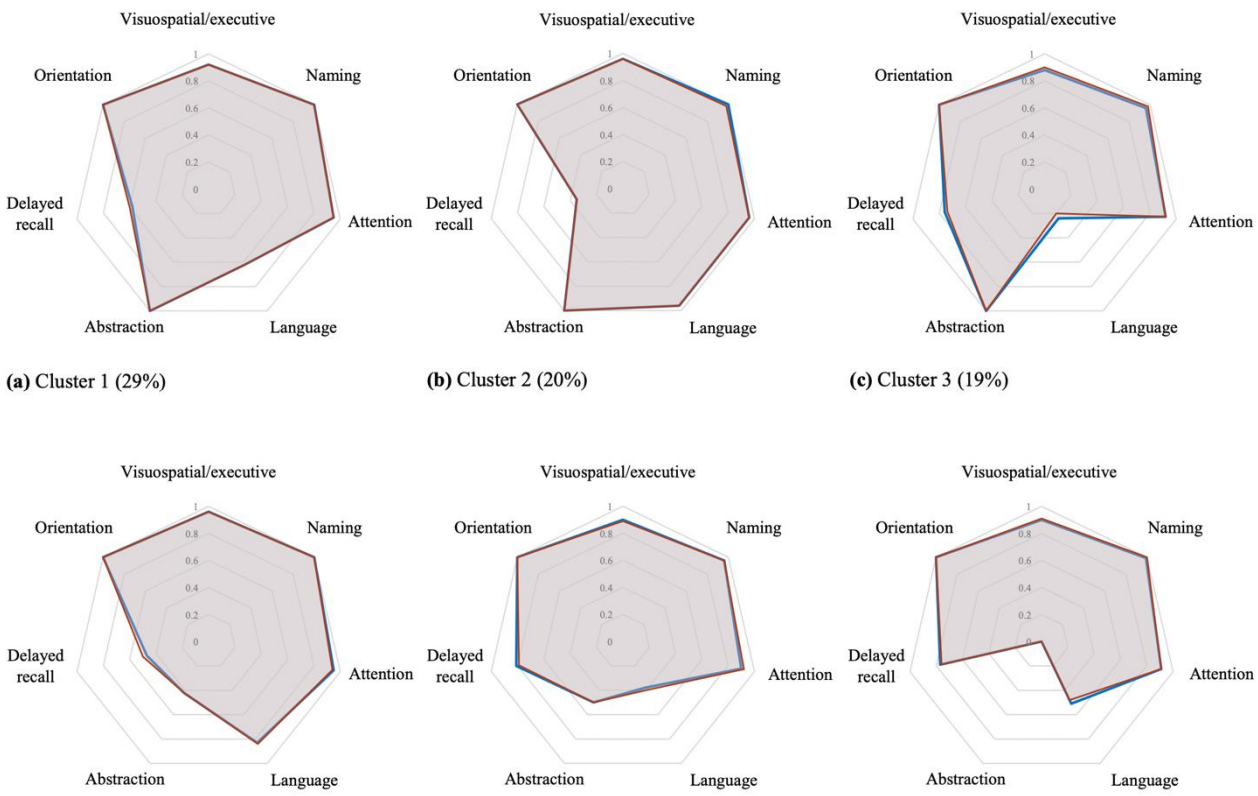


Figure 6. Unsupervised machine learning demonstrated six independent clusters among aging (≥60 years old) Thai population with borderline mild cognitive impairment.[32]

CHAPTER 3

STUDY DESIGN AND METHODOLOGY

Design

A retrospective descriptive study of aging Thai PLWH

Methodology

Population

The study used the secondary data from participants who were enrolled in the HIV Netherlands Australia Thailand Research Collaboration (HIV-NAT) 207 ("Aging study"). The HIV-NAT 207 study was a cross-sectional study to determine the prevalence and compare multiple health issues such as osteoporosis, vitamin D deficiency, cardiovascular risks, depression, and neurocognitive impairment between aging PLWH and age sex-matched HIV-uninfected controls. The study enrolled two groups of participants between 2015 and 2017 for a total of 509 participants: PLHW from the ongoing prospective HIV-NAT 006 study (n=358) and the newly enrolled HIV-uninfected controls (n=151).[33-36] The HIV-NAT 207 study measured several aged-related factors of interest such as blood pressure, waist circumference, height, weight, bioelectrical impedance analysis (BIA), questionnaires (assessing fracture risks, physical activities, cardiovascular risks, depression and anxiety, quality of life, activities of daily living, and nutrition), bone mineral density of hip and spine, metabolic and hormonal laboratories, transient elastography, coronary artery calcium (CAC) score, and transthoracic echocardiogram. Importantly, all aging PLWH who were enrolled in the HIV-NAT 207 study were the participants who had already been enrolled in the prospective and ongoing HIV-NAT 006 study. The HIV-NAT 006 study is

an ongoing, prospective, clinic-based cohort that enrolled adults living with HIV (aged ≥18 years) since 1996 (NCT00411983) and have enrolled over 2000 participants as of 2021.[37-41] Therefore, all participants in our study had baseline data prior to and the current data after the MoCA in the HIV-NAT 207 study, conducted between 2015 and 2017. Among many laboratory parameters and other investigations performed in the HIV-NAT 207 study was the Thai-validated version of MoCA, which was used to assess cognitive performance.

Target Population

Aging PLWH who were enrolled in the HIV-NAT 207 study

Control Population

Because the study was a descriptive study of aging Thai PLWH, there was no control population.

Approach to participants

Because the study used secondary data from the already completed the HIV-NAT 207 study, there was no approach to participants.

Inclusion criteria

• Aging PLWH, defined by PLWH who aged ≥50 years old, who were enrolled in the HIV-NAT 207 study

• Completed the MoCA

Exclusion criteria

• Participants who were previously diagnosed with dementia

Sample size calculation

Unsupervised ML uses data-driven analysis to determine hidden patterns in the dataset without the absolute ground truth. Therefore, the sample size cannot be

calculated since we plan to integrate this method to determine the solution to our primary objective. All eligible aging PLWH who were enrolled in the HIV-NAT 207 study were included in our study. Clustering validation was conducted after the cluster analysis to determine the performance of the model.

Informed consent process

All participants in the HIV-NAT 207 study provided their consent prior to entering the study. Because the study used secondary data from the already completed the HIV-NAT 207 study, the informed consent for this study was waived.

Data Collection

Secondary data from the HIV-NAT 207 study was extracted and used for the analysis. The data necessary for the primary objective of the study was the participants' scores from the MoCA. This included the total score of each of the seven cognitive domains in the MoCA: visuospatial/executive function, naming, attention, language, abstraction, delayed recall, and orientation.

Demographic data comprised age, gender, education levels, marital status, body mass index (BMI), smoking, alcohol consumption, hypertension, diabetes mellitus, dyslipidemia, and depression. BMI was calculated as weight in kilograms divided by the square of the height in meters and categorized based on the WHO classifications for Asian populations.[42] Hypertension was defined as systolic blood pressure of ≥140 mmHg or diastolic blood pressure of ≥90 mmHg on two or more clinic visits or reported onset of hypertension and initiation of antihypertensive therapy by a physician. Diabetes mellitus was defined as fasting plasma glucose of ≥126mg/dl on two consecutive study visits or reported onset of diabetes and initiation of antidiabetic therapy by a physician. Depression was evaluated by a 15-

item depression screening test developed and validated by the Department of Mental Health, Ministry of Public Health of Thailand, with a score of >6/15 indicating clinical depression. HIV-related factors included the CDC classification system, nadir CD4 cell counts, duration of HIV diagnosis and ART initiation, history of efavirenz use, current CD4 cell counts, and plasma HIV RNA levels.

Data Analysis and Statistics

Demographic data

Demographic data was summarized as number and percentage for categorical variables and as median and interquartile for continuous variables. To determine the differences in demographic features between clusters of aging Thai PLWH with borderline cognitive impairment, Pearson's chi-square, Fisher's exact test, and Kruskal–Wallis test were used as appropriated.

Unsupervised machine learning algorithm

According to our primary objective of identifying hidden patterns or intrinsic structures in the data, or "cluster", unsupervised ML is the type of ML used in this study. K-means clustering was used as the unsupervised ML algorithm of choice to perform hard clustering, where each data point belongs to only one cluster. K-means clustering partitions the given data into user-specified *k* number of mutually exclusive clusters, with each cluster having a cluster center called centroid. The distance from that point to the cluster's center determines how well a point fits into each cluster.[25 43] Given *k*, the k-means algorithm works as follows:

Step 1: Choose *k* data points to be the initial centroids (cluster centers).

Step 2: Assign each data point to the closest centroids.

Step 3: Recompute the centroids using the current cluster memberships.

Step 4: Repeat step 2 and 3 until convergence criteria is met.

The input variables for cluster analysis were the total score of each of the seven cognitive domains in the MoCA:

- Visuospatial/executive function (maximum score: 5)
- Naming (3)
- Attention (6)
- Language (3)
- Abstraction (2)
- Delayed recall (5)
- Orientation (6)

Clustering validation

Unlike supervised ML, there is no ground truth to unsupervised ML, and thus identify the most appropriate number of clusters remains one of the most perplexing issues. There are several methods to determine the appropriate number of clusters can be determined, although there is currently no universal standard recommendation. Several methods were used to determine the optimal number of clusters in our study to be most vigilant on this important issue. These included (1) elbow method, (2) Silhouette coefficient, (3) Calinski-Harabasz Index, (4) Davies-Bouldin Index, and split data method.[44-47]

The elbow method is among the most popular methods for determining the optimal number of clusters. The method is based on calculating the Within-Cluster-Sum of Squared Errors (WSS) for the different number of clusters (k) and selecting the "elbow point", which is the point where the variation changes began to diminish. The elbow point is the number of clusters we can use for our clustering algorithm.[44]

The Silhouette Coefficient calculates the smallest average distance of point *i* to all points in other clusters (distance of points of different clusters) and the average distance of point *i* from all points in the same cluster (closeness of points in the same cluster). Therefore, the model provides the information on whether the individual points are correctly assigned to their clusters – i.e. the point would be better off assigning to the other clusters if the Silhouette Coefficient is closer to -1, and the point belongs to the 'correct cluster if the Silhouette Coefficient is closer to 1.[45]

The Calinski-Harabasz index, also known as the Variance Ratio Criterion, is the ratio of between-clusters dispersion and inter-cluster dispersion for all clusters; the higher the score, the better the performances.[46]

This Davies-Bouldin index signifies the average 'similarity' between clusters, where the similarity is a measure that compares the distance between clusters with the size of the clusters themselves. A lower Davies-Bouldin index relates to a model with better separation between the clusters.[47]

Lastly, the application of split data for training and validating used in supervised learning was applied to the cluster's outcome.[48] After the optimal number of clusters was determined using the abovementioned methods, the dataset was split into 1:1 ratio and clustered independently using the chosen number of clusters. The consistency between datasets would suggest the validity of the model.

Follow-up data

Although not among the primary proposes of this cross-sectional description study which determined to demonstrate the capability of unsupervised ML in exploring clusters of aging PLWH with borderline cognitive impairment, longitudinal information on plasma HIV RNA was available by matching the data from the HIV-NAT

207 to the longitudinal HIV-NAT 006 database. The incidence of detectable plasma HIV RNA (≥40 copies/mL) was calculated for each cluster and demonstrated using the Kaplan-Meier estimate. Participants without detectable plasma HIV RNA were censored on the last day they had plasma HIV RNA levels. Person-years were calculated from enrollment of the HIV-NAT 207 to the first visit of detectable plasma HIV RNA for participants who had detectable plasma HIV RNA and the last visit with plasma HIV RNA result for participants who did not have detectable plasma HIV RNA.

All statistical analysis were performed using Stata/SE 17 (StataCorp LP, College Station, TX, USA) except for the unsupervised ML algorithm, which was performed using R (RStudio, Boston, MA, USA).

CHAPTER 4

RESULTS

Characteristics of all enrolled participants

Among 340 PLWH aged ≥50 years who completed the MoCA in the HIV-NAT 207 study, 177 (52.1%) scored between 23 and 27 (median [IQR] 25 [24 – 26]) and were included in the analysis. Median (IQR) age was 54 (51-58) years, 118 (66.7%) were male, and 149 (84.2%) had >6 years of education. For comorbidities, 67 (37.9%) have ever smoked, 32 (18.1%) have ever consumed alcohol, 65 (36.7%) had hypertension, 28 (15.8%) had diabetes mellitus, 65 (37.8%) had LDL-cholesterol of >130 mg/dL, 90 (50.9%) had hypertriglyceridemia, and 38 (21.5%) had depression **(Table 3)**.

The median duration of HIV diagnosis and ART initiation were 18.2 (15.2 – 20.9) and 16.3 (13.6 – 19.1) years, respectively. 110 (62.2%) had history of EFV use. Median current CD4 counts were 620 (489 – 795) cells/μL and 170 (96.1%) had plasma HIV RNA <40 copies/mL.

Table 3. Characteristics of 177 participants

Characteristics	**N**	**%**
Gender		
Female	59	33.3
Male	118	66.7
Age, y[ǂ]	54 (51 – 58)	
Education		
Primary school or less	28	15.8
Secondary school	83	46.9
Bachelor's degree or more	66	37.3
Marital status		
Married	71	40.6
Single/separated/widowed	104	59.4
Body mass index (kg/m^2)		
Underweight (<18.5)	18	10.2
Ideal (18.5 – 22.9)	68	38.6
Overweight (23 – 27.5)	76	43.2
Obese (>27.5)	14	8.0
Smoking status		
Current	27	15.3
Former	40	22.6
Never	110	62.1
Alcohol consumption status		
Current	14	7.9
Former	18	10.2
Never	145	81.9
Hypertension	65	36.7
Diabetes mellitus	28	15.8
LDL-cholesterol >130 mg/dL	65	37.8

Triglyceride >150 mg/dL	90	50.9
Depression	38	21.5
HIV-related factors		
CDC classification		
A	75	42.4
B	72	40.7
C	30	16.9
Nadir CD4 counts, cells/µl[ǂ]	191 (74 – 266)	
Duration of HIV diagnosis, y[ǂ]	18.2 (15.2 – 20.9)	
Duration of ART initiation, y[ǂ]	16.3 (13.6 – 19.1)	
History of EFV use	110	62.2
Current CD4 counts, cells/µl[ǂ]	620 (489 – 795)	
Plasma HIV RNA <40 copies/mL	170	96.1
MoCA score[ǂ]	25 (24 –26)	

Abbreviations: ART, antiretroviral therapy; EFV, efavirenz; MoCA, Montreal Cognitive Assessment

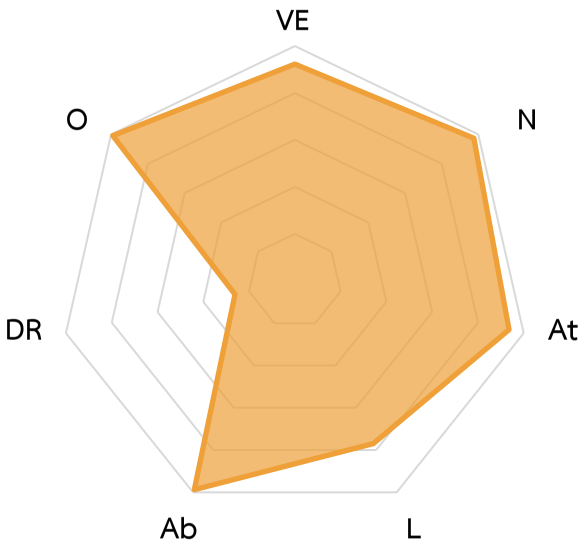
[ǂ] Continuous variables are reported as median (interquartile range)

Clustering outcomes

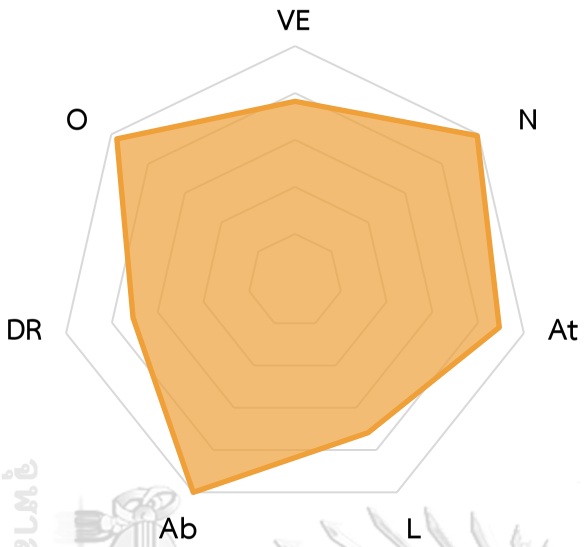
Unsupervised ML demonstrated 5 clusters with good consistency of borderline cognitive impairment among aging Thai PLWH **(Figure 7,** cluster validation results were demonstrated in **Figure 8 and 9)**: 22.0% cluster 1 (marked memory with mild language impairment), 25.4% cluster 2 (mild visuospatial/executive function-language-memory impairment), 19.2% cluster 3 (moderate abstraction with mild visuospatial/executive function-language-memory impairment), 18.6% cluster 4 (marked language with mild memory impairment), 14.7% cluster 5 (marked language-abstraction impairment).

Differences in characteristics between clusters

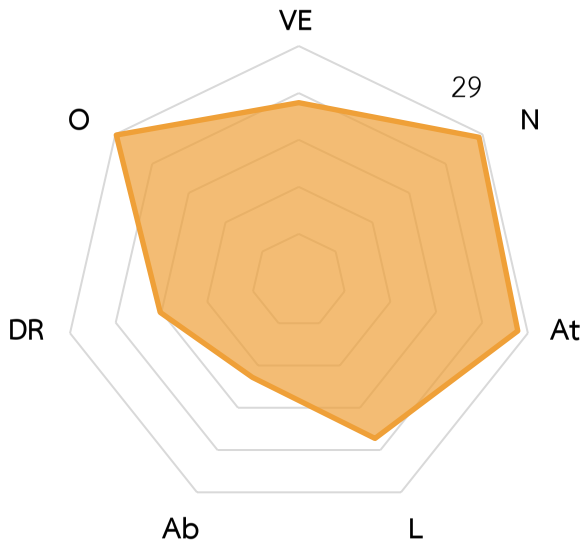
Characteristics of participants in each cluster was demonstrated in **Table 4**. There were significant differences in the proportion of participants with education of primary school or less between clusters (p=0.012): cluster 2 (24.4%) and 4 (27.3%) had the highest proportion, and cluster 1 (2.6%) had the lowest proportion. No significant differences in other participant's characteristics, including HIV-related factors, between clusters were shown. There were significant differences in the median MoCA score between clusters (p<0.001): 26 for cluster 2; 25 for cluster 1; and 24 for cluster 3 – 5.



a) Cluster 1 (22.0%): Marked memory with mild language impairment (DR+++, L+)



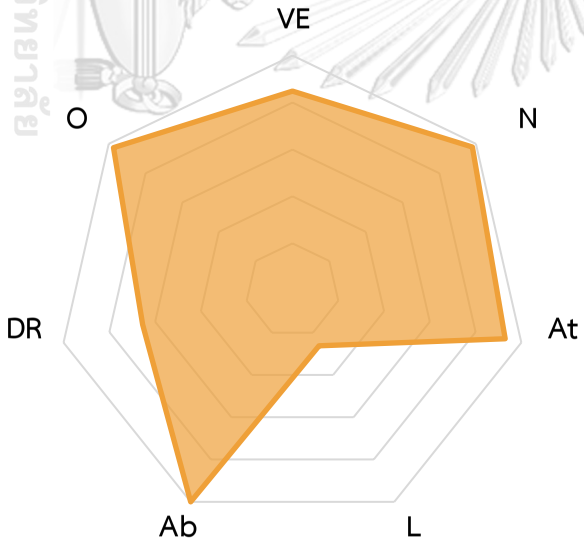
b) Cluster 2 (25.4%): Mild visuospatial/executive function-language-memory impairment (VE+, L+, DR+)



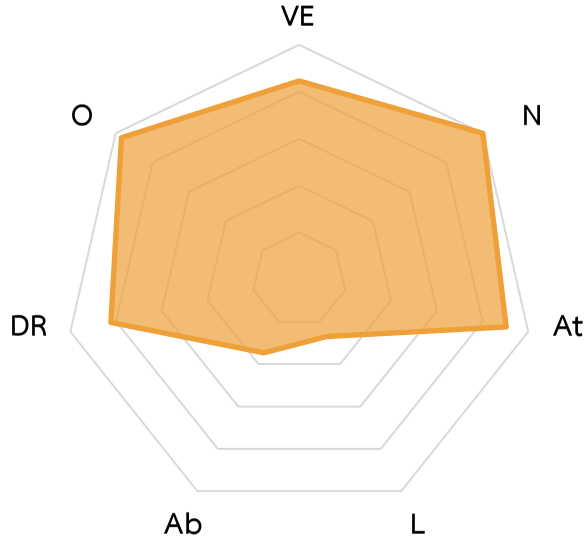
c) Cluster 3 (19.2%): Moderate abstraction with mild visuospatial/executive function-language-memory impairment (Ab++, VE+, L+, DR+)

Figure 7. Unsupervised machine learning demonstrated 5 clusters (a-e) of borderline cognitive impairment among aging Thai PLWH. Abbreviations: VE, visuospatial/executive; N, naming; At, attention; L, language; Ab, abstraction; DR, delayed recall; O, orientation

* +, ++, and +++ represented mild, moderate, and marked impairment, respectively



d) Cluster 4 (18.6%): Marked language with mild memory impairment (L+++, DR+)



e) Cluster 5 (14.7%): Marked language-abstraction impairment (L+++, Ab+++)

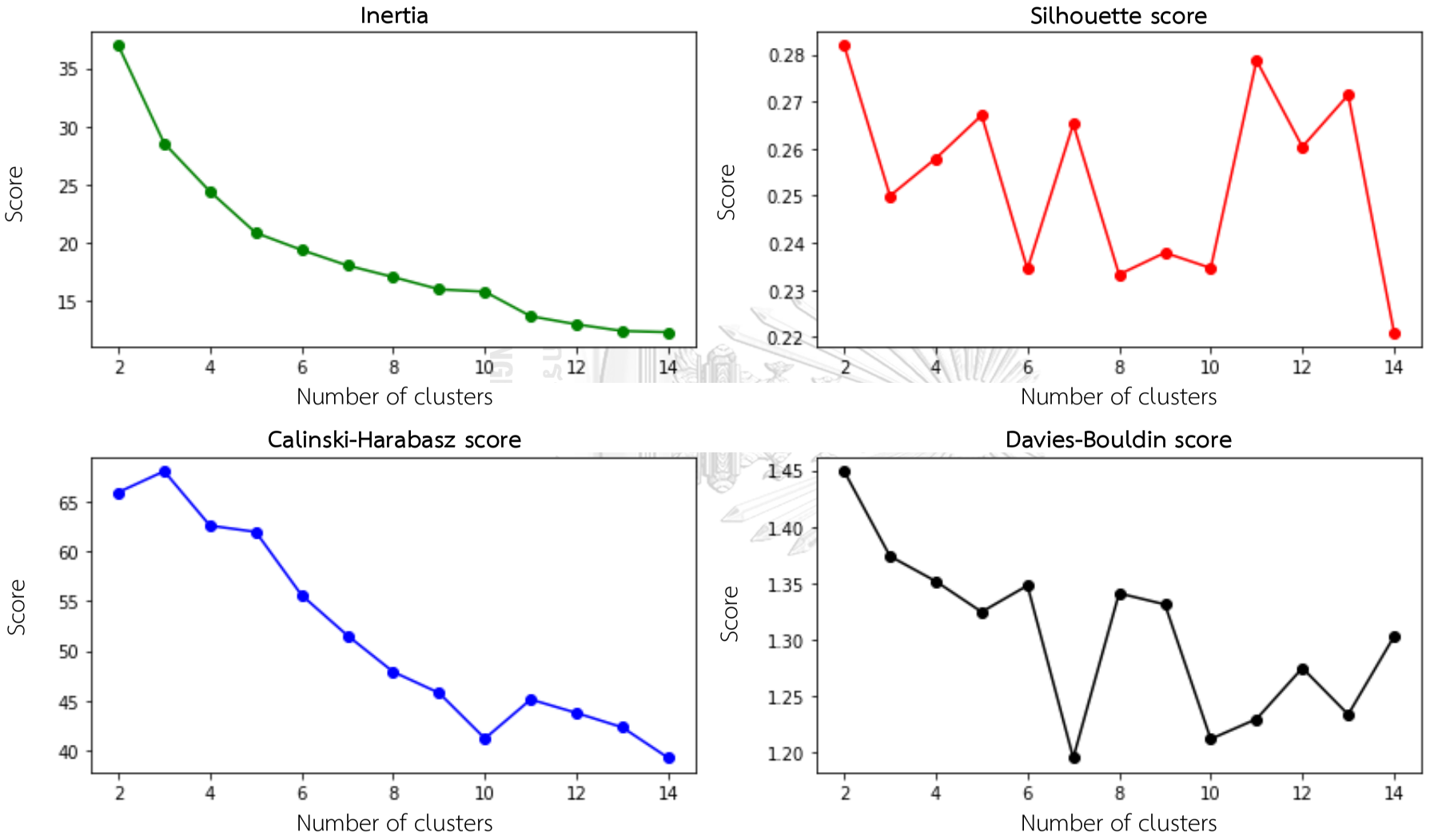
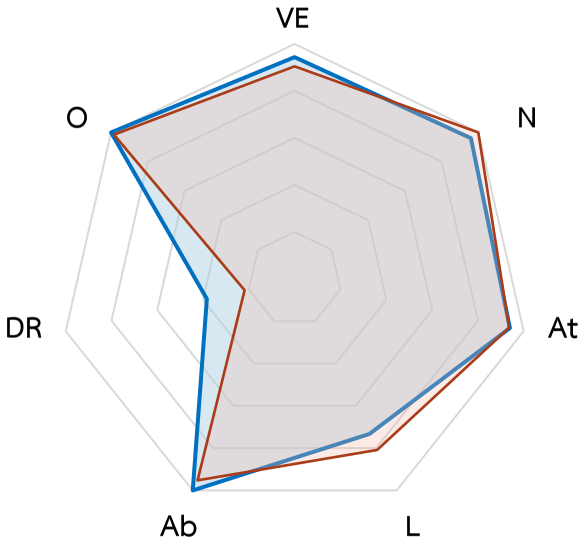
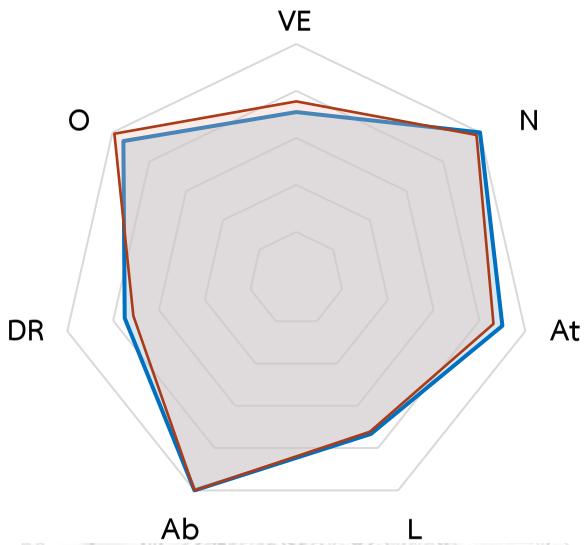


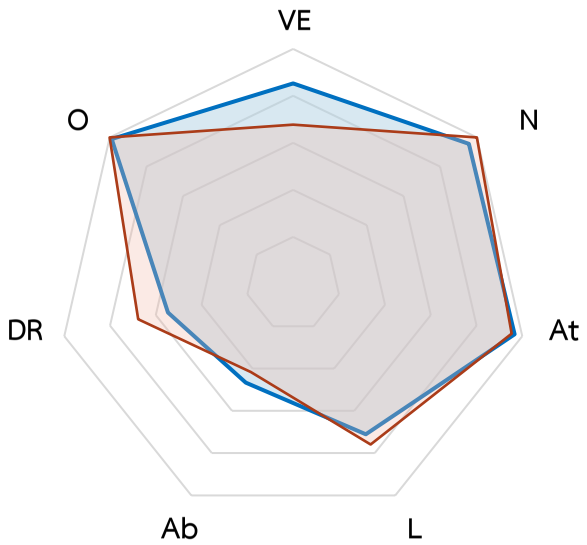
Figure 8. Clusters evaluation using the elbow method on the inertia curve, silhouette score, Calinski-Harabasz score, Davies Bouldin score.



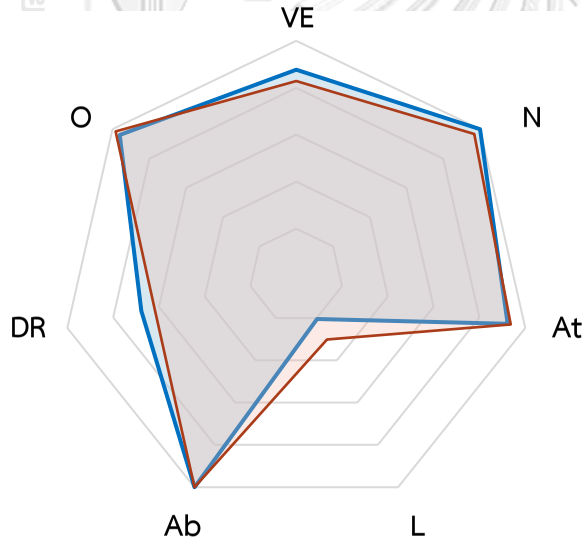
Cluster 1



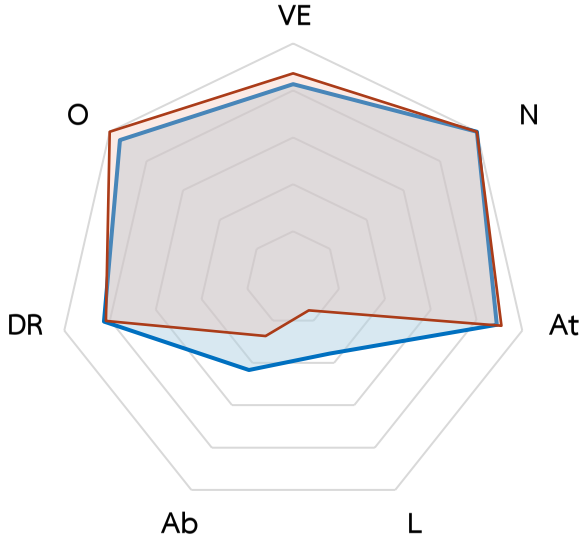
Cluster 2



Cluster 3



Cluster 4



Cluster 5

Figure 9. Cluster evaluation using split data method. Data was split into two (1:1) and clustered independently to examine the consistency of the unsupervised machine learning model, with blue and red line representing the first and second half of the data, respectively.

Abbreviations: VE, visuospatial/executive; N, naming; At, attention; L, language; Ab, abstraction; DR, delayed recall; O, orientation

Table 4. Characteristics of participants categorized by clusters

Characteristics	Cluster 1 N=39		Cluster 2 N=45		Cluster 3 N=34		Cluster 4 N=33		Cluster 5 N=26		p-value
	n	%	n	%	n	%	n	%	n	%	
Gender											0.826[1]
Female	12	30.8	17	37.8	13	38.2	9	27.3	8	30.8	
Male	27	69.2	28	62.2	21	61.8	24	72.7	18	69.2	
Age, y[ǂ]	53 (51 –59)		54 (51 – 57)		54.5 (51 – 57)		56 (53 – 61)		53 (52 – 55)		0.223[3]
Education											0.012[1]
Primary school or less	1	2.6	11	24.4	5	14.7	9	27.3	2	7.7	
Secondary school	15	38.5	21	46.7	20	58.8	13	39.4	14	53.8	
Bachelor degree or more	23	59.0	13	28.9	9	26.5	11	33.3	10	38.5	
Marital status											0.660[1]
Married	14	35.9	22	48.9	11	33.3	14	42.4	10	40.0	
Single/separated/widowed	25	64.1	23	51.1	22	66.7	19	57.6	15	60.0	
Body mass index (kg/m2)											0.418[2]
Underweight (<18.5)	4	10.3	4	9.1	4	11.8	2	6.1	4	15.4	
Ideal (18.5 – 22.9)	9	23.1	21	47.7	17	50.0	13	39.4	8	30.8	
Overweight (23 – 27.5)	22	56.4	16	36.4	11	32.4	14	42.4	13	50.0	
Obese (>27.5)	4	10.3	3	6.8	2	5.9	4	12.1	1	3.8	
Smoking status											0.900[1]
Current	5	12.8	5	11.1	5	14.7	7	21.2	5	19.2	
Former	9	23.1	11	24.4	7	20.6	9	27.3	4	15.4	
Never	25	64.1	29	64.4	22	64.7	17	51.5	17	65.4	

Alcohol consumption status											0.786[2]
Current	2	5.1	5	11.1	3	8.8	2	6.1	2	7.7	
Former	4	10.3	4	8.9	3	8.8	6	18.2	1	3.8	
Never	33	84.6	36	80.0	28	82.4	25	75.8	23	88.5	
Hypertension	21	53.8	11	24.4	12	35.3	12	36.4	9	34.6	0.094[1]
Diabetes mellitus	8	20.5	5	11.1	9	26.5	2	6.1	4	15.4	0.155[1]
LDL >130 mg/dL	16	43.2	19	43.2	11	34.4	12	36.4	7	26.9	0.644[1]
Triglyceride >150 mg/dL	23	59.0	24	53.3	16	47.1	16	48.5	11	42.3	0.705[1]
Depression	8	20.5	8	17.8	5	14.7	13	39.4	4	15.4	0.086[1]
HIV-related factors											
CDC classification											0.545[1]
A	20	51.3	18	40.0	11	32.4	15	45.5	11	42.3	
B	12	30.8	22	48.9	14	41.2	12	36.4	12	46.2	
C	7	17.9	5	11.1	9	26.5	6	18.2	3	11.5	
Nadir CD4 counts, cells/μl[ǂ]	206 (97 – 315)		191 (114 – 245)		123 (57 – 244)		232 (58 – 289)		155.5 (49 – 233)		0.315[3]
Duration of HIV diagnosis, y[ǂ]	17.9 (13.9 – 20.1)		19.2 (16.0 – 21.9)		18.6 (15.6 – 20.9)		17.6 (15.7 – 19.3)		19.7 (14.0 – 22.1)		0.625[3]
Duration of ART initiation, y[ǂ]	16.8 (13.9 – 19.1)		16.4 (12.6 – 19.4)		15.5 (11.8 – 18.7)		16.2 (14.6 – 18.1)		16.2 (14.4 – 19.7)		0.875[3]
History of EFV use	28	71.8	31	68.9	16	47.1	19	57.6	16	61.5	0.199[1]
Current CD4 counts, cells/μl[ǂ]	627 (441 – 948)		644 (496 – 803)		573 (442 – 753)		697 (505 – 871)		604 (511 – 667)		0.544[3]
HIV RNA <40 copies/mL	37	94.9	44	97.8	32	94.1	32	97.0	25	96.2	0.924[2]
MoCA score[ǂ]	25 (24 – 25)		26 (25 – 27)		24 (24 – 25)		24 (24 – 26)		24 (23 – 25)		0.0001[3]

Abbreviations: ART, antiretroviral therapy; EFV, efavirenz; MoCA, Montreal Cognitive Assessment

[ǂ] Continuous variables are reported as median (interquartile range)

[1] Pearson's chi-square

[2] Fisher's exact test

[3] Kruskal–Wallis test

Clusters and virological outcomes

Among 177 participants, 167 (94.4%) had plasma HIV RNA follow-up data. A total of 16 (9.6%) aging PLWH with borderline cognitive impairment had incident plasma HIV RNA detection (≥40 copies/mL), accounting for an overall incidence of 22.2 per 1000 person-years (95%CI 13.6-36.2) over 720.7 person-years of follow-up. The incidence of plasma HIV RNA detection was 31.6, 16.3, 28.8, 32.2, and 0 per 1000 person-years for cluster 1-5, respectively **(Figure 10)**.

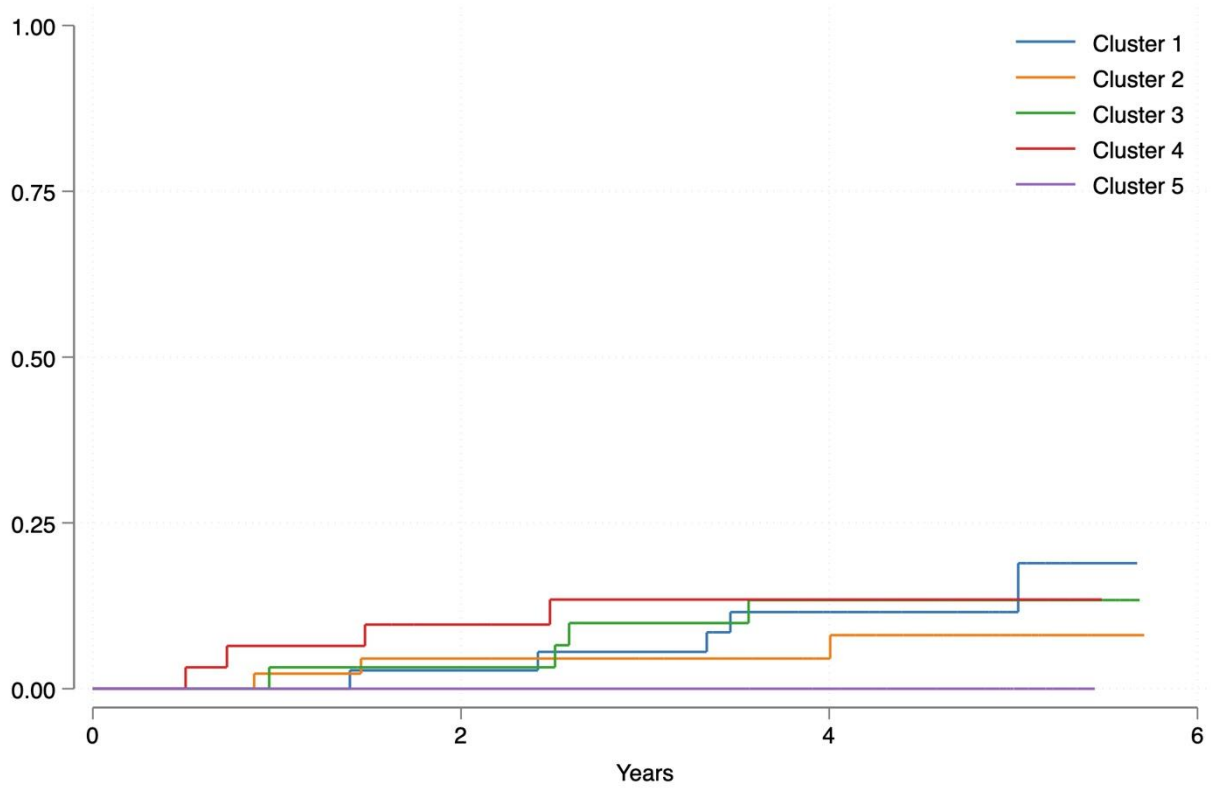


Figure 10. Kaplan-Meier estimates of having plasma HIV RNA detection (≥40 copies/mL) categorized by clusters. Over a total of 720.7 person-years of follow-up, the incidence of plasma HIV RNA detection was 31.6, 16.3, 28.8, 32.2, and 0 per 1000 person-years for cluster 1-5, respectively.

CHAPTER 5

DISCUSSION

Unsupervised ML demonstrated five clusters with good consistency surrounding visuospatial/executive, abstraction, language, and memory function among aging Thai PLWH with borderline cognitive impairment; all had some degree of language impairment.

In contrast to the previous studies that applied ML to cognitive impairment,[28-31 49] we did not use a new diagnostic tool, advanced laboratory, or advanced imaging; and opted to use the MoCA exclusively as our cluster variables. This is beneficial to apply the method in many healthcare services, particularly in resource-limited settings where advanced investigations may not be feasible. In addition, cognitive assessment tools that are not commonly used would be difficult to perform and integrate into many healthcare settings. By introducing a novel approach to interpreting the data from the already well-recognized assessment tool, our study provided an effective and uncomplicated way for healthcare workers to implement this approach in healthcare settings

All five clusters had some degree of language impairment. It is important to note that this might not automatically point to abnormalities of the cerebral cortex, the traditional anatomical localization of the physiology of language.[50 51] Language assessment in the MoCA consisted of sentence repetition and verbal fluency. In addition to understanding and speaking the language, the ability to repeat sentences requires core components of working memory, such as holding information for a brief period and computing syntactic structures to rehearse information successfully.[52] The ability to initiate the word in verbal fluency also requires components of executive function.[53] Therefore, the ability to perform both tasks of the MoCA's language assessment may correlate with individual differences in working memory and

executive function capacity, localized at the subcortical regions where they are predominantly affected in PLWH.[54]

All impairments surrounded four out of seven cognitive domains tested in the MoCA – visuospatial/executive, abstraction, language, and memory function. Since all enrolled participants scored near the cut-off point of the MoCA, we hypothesized that this represents the cognitive domains that are impaired at the early stage of cognitive impairment. It is most likely that these cognitive domains require a relatively higher brain function network than the remaining three cognitive domains that were not affected. Alternatively, secondly, these findings may represent a bias of measurement. Although visuospatial/executive function was scored collectively, it was tested thoroughly using three tests – alternating trail making, cube drawing, and clock drawing test. Language assessment in the MoCA requires multiple brain regions, as previously discussed. The score for word recall did not differentiate between those who can and cannot recall after the cue. Comparing to our previous work using the same clustering methods on the general aging population attending the King Chulalongkorn Memorial Hospital, Bangkok, Thailand, demonstrated six clusters of aging Thai population, two clusters did not have language impairment despite the study's median age was higher than the current study by more than a decade (median age 66.9 years).[32]

Although significant differences in the median total MoCA score between clusters were demonstrated, all participants scored within a narrow range of 23-27. Notably, the median MoCA scores were identical among clusters 3 – 5, confirming that different phenotypes exist among those with a similar score. Furthermore, the median score of the two clusters was within the traditional cut-off point of 25, suggesting that the traditional binary outcomes of cognitive assessment would have missed important in-depth information about the patients.

In relation to the four traditional subtypes of MCI, unsupervised ML demonstrated four amnestic MCI-multiple domains (cluster 1 – 4) and a group of non-amnestic MCI-multiple domains (cluster 5). There are currently no US Food and Drug Administration (FDA)-approved medications for the treatment of MCI. Numerous control trials have been conducted, but none have shown effectiveness in delaying the progression from MCI to dementia.[1] Therefore, the current recommendations suggest that clinicians counsel patients and families that no pharmacologic agents are shown to have symptomatic cognitive benefit in MCI.[7] Based on our findings that different phenotypes exist among those with a similar score from a cognitive assessment, the lack of effective pharmacological interventions may be partly due to the previously undiscovered heterogeneity of the population.

Previous reports have demonstrated that cognitive impairment among PLWH can affect their ability to adequately adhere to ART,[55 56] which may lead to detectable plasma HIV RNA. By linking the data from our aging PLWH with borderline cognitive impairment to the longitudinal HIV-NAT 006 study, we were able to obtain the follow-up data of 167 (94.4%) participants. Although there were differences between clusters, plasma HIV RNA detection incidence was low across all clusters (ranged 0 – 32.3 per 1000 person-years). It is crucial to note that all participants are at most in their early stage of cognitive impairment and, therefore, might not wholly reflect the previous reports on the effect of cognitive impairment on ART adherence. However, the interesting information from our findings was that the only cluster that did not show incident detectable HIV RNA was cluster 5 – the only non-amnestic cluster, in line with previous reports on the association between memory function and medication management skills.[57-59]

Certain limitations need to be considered in our study. Because our study was conducted in a research center in a large city, there may be selection bias, as can be seen in the disproportion of individuals with an education of >6 years. Secondly, we cannot conclude the extent of the clinical significance of these clusters because of

the cross-sectional study design fashion. However, we were able to show some light on the potential differences in virological outcomes between clusters by linking to the existing longitudinal data. Thirdly, the sample size was small, which can affect the consistency of the clusters, as can be best viewed by the split data method, and the ability to perform multivariable analysis to explore associated factors. It is also important to note that our participants were mostly virologically well-controlled aging PLWH who mostly came for a routine follow-up rather than presenting with neurological issues. Finally, because the data collection of the MoCA score was in the form of total score per each of the 7 cognitive domains assessed, this limited the potential to regroup the score of each test within those cognitive domains to further explore tests that may have overlapping cognitive domains assessment values. Regardless, we were able to accomplish our objective of determining clusters of aging Thai PLWH with borderline cognitive impairment.

Our study demonstrated that heterogeneity exists within those who score similarly in the cognitive assessment. The findings can be used to tailor the selection of cognitive domain assessment in the usually busy clinic settings. However, among the most critical question moving forward is whether these clusters have any clinical significance. Therefore, a longitudinal study is warranted to identify the effect on HIV care (ART adherence and viral rebound incidence), cognitive progression, quality of life, and mortality. Studies using unsupervised ML in different population is also beneficial for exploring the difference in the clusters. Finally, inputting additional data on clinical history/physical examination and other investigations such as brain imaging and biomarkers may improve the cluster outcomes for settings with available resources.

In conclusions, unsupervised ML demonstrated five clusters among aging Thai PLWH with borderline cognitive impairment; all had some degree of language impairment and multiple domains involvement. A longitudinal study is warranted to identify differences in clinical significance and prognosis between each cluster.

REFERENCES

1. Petersen RC. Mild Cognitive Impairment. Continuum (Minneapolis, Minn) 2016;**22**(2 Dementia):404-18.
2. National Statistical Office of Thailand. Demography Population and Housing statistics. 2020 [29 July 2021]. Available from: http://statbbi.nso.go.th/staticreport/page/sector/th/01.aspx.
3. Department of Older Persons (DOP). Situation of the Thai Elderly. Bangkok, Thailand: Foundation of Thai Gerontology Research and Development Institute (TGRI); 2016 [29 July 2021]. Available from: https://thaitgri.org/?wpdmpro=situation-of-the-thai-eldery-2019.
4. World Health Organization. The top 10 causes of death. 2020 [29 July 2021]. Available from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
5. Numbers K, Brodaty H. The effects of the COVID-19 pandemic on people with dementia. Nature reviews Neurology 2021;**17**(2):69-70.
6. Gauthier S, Reisberg B, Zaudig M, et al. Mild cognitive impairment. Lancet (London, England) 2006;**367**(9518):1262-70.
7. Petersen RC, Lopez O, Armstrong MJ, et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology 2018;**90**(3):126-35.
8. Valcour V, Chalermchai T, Sailasuta N, et al. Central nervous system viral invasion and inflammation during acute HIV infection. The Journal of infectious diseases 2012;**206**(2):275-82.
9. Osborne O, Peyravian N, Nair M, et al. The Paradox of HIV Blood-Brain Barrier Penetrance and Antiretroviral Drug Delivery Deficiencies. Trends in neurosciences 2020;**43**(9):695-708.
10. Neurodevelopmental delay in pediatric HIV/AIDS: current perspectives; 2016.
11. Saylor D, Dickens AM, Sacktor N, et al. HIV-associated neurocognitive disorder--

pathogenesis and prospects for treatment. Nature reviews Neurology 2016;**12**(4):234-48.

12. Skavronskaya L, Moyle B, Scott N. The Experience of Novelty and the Novelty of Experience. Frontiers in psychology 2020;**11**:322.
13. Harada CN, Natelson Love MC, Triebel KL. Normal cognitive aging. Clinics in geriatric medicine 2013;**29**(4):737-52.
14. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. Archives of neurology 1999;**56**(3):303-8.
15. Ferman TJ, Smith GE, Kantarci K, et al. Nonamnestic mild cognitive impairment progresses to dementia with Lewy bodies. Neurology 2013;**81**(23):2032-8.
16. Vega JN, Newhouse PA. Mild cognitive impairment: diagnosis, longitudinal course, and emerging treatments. Current psychiatry reports 2014;**16**(10):490.
17. Bäckman L, Jones S, Berger AK, et al. Multiple cognitive deficits during the transition to Alzheimer's disease. Journal of internal medicine 2004;**256**(3):195-204.
18. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society 2005;**53**(4):695-9.
19. Milani SA, Marsiske M, Cottler LB, et al. Optimal cutoffs for the Montreal Cognitive Assessment vary by race and ethnicity. Alzheimers Dement (Amst) 2018;**10**:773-81.
20. Tangwongchai S, Charernboon T, Phanasathit M, et al. *The validity of thai version of the montreal cognitive assessment (MoCA-T)*, 2009.
21. Winston A, Spudich S. Cognitive disorders in people living with HIV. The lancet HIV 2020;**7**(7):e504-e13.
22. Antinori A, Arendt G, Becker JT, et al. Updated research nosology for HIV-associated neurocognitive disorders. Neurology 2007;**69**(18):1789-99.
23. Papadokostaki K, Mastorakis G, Panagiotakis S, et al. Handling Big Data in the Era of Internet of Things (IoT). In: Mavromoustakis CX, Mastorakis G, Dobre C, eds. Advances in Mobile Cloud Computing and Big Data in the 5G Era. Cham: Springer International Publishing, 2017:3-22.
24. Myszczynska MA, Ojamies PN, Lacoste AMB, et al. Applications of machine learning

to diagnosis and treatment of neurodegenerative diseases. Nature Reviews Neurology 2020;**16**(8):440-56.

25. MathWorks. Introducing Machine Learning., 2016.
26. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;**542**(7639):115-18.
27. Anderson JP, Parikh JR, Shenfeld DK, et al. Reverse Engineering and Evaluation of Prediction Models for Progression to Type 2 Diabetes: An Application of Machine Learning Using Electronic Health Records. Journal of diabetes science and technology 2015;**10**(1):6-18.
28. Park JH. Machine-Learning Algorithms Based on Screening Tests for Mild Cognitive Impairment. American journal of Alzheimer's disease and other dementias 2020;**35**:1533317520927163.
29. Yim D, Yeo TY, Park MH. Mild cognitive impairment, dementia, and cognitive dysfunction screening using machine learning. The Journal of international medical research 2020;**48**(7):300060520936881.
30. Chiu PY, Tang H, Wei CY, et al. NMD-12: A new machine-learning derived screening instrument to detect mild cognitive impairment and dementia. PloS one 2019;**14**(3):e0213430.
31. Escudero J, Zajicek JP, Ifeachor E. Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference 2011;**2011**:6470-3.
32. Hiransuthikul A, Chunharas C, Thongkam A, et al. Determining In-Depth Information of Mild Cognitive Impairment: Clustering Using Unsupervised Machine Learning (1352). Neurology 2021;**96**(15 Supplement):1352.
33. Apornpong T, Han WM, Chattranukulchai P, et al. Higher Proportion of Abnormal Nutritional Status Among Well-Suppressed HIV-Infected Elderly Asians Compared to HIV-Negative Individuals. AIDS Res Hum Retroviruses 2020;**36**(7):590-96.
34. Chattranukulchai P, Thimaporn W, Siwamogsatham S, et al. Echocardiographic Findings Among Virally Suppressed HIV-Infected Aging Asians Compared with

HIV-Negative Individuals. J Acquir Immune Defic Syndr 2020;**85**(3):379-86.

35. Hiransuthikul A, Chutinet A, Sakulrak S, et al. Short Communication: Carotid Intima-Media Thickness Is Not Associated with Neurocognitive Impairment Among People Older than 50 Years With and Without HIV Infection from Thailand. AIDS research and human retroviruses 2019;**35**(11-12):1170-73.
36. Thet D, Siritientong T, Sangarlangkarn A, et al. Deterioration of Nutritional Status and Its Negative Association with Depression Among Older HIV-Infected Asian Population: A Four-Year Longitudinal Study. AIDS Res Hum Retroviruses 2021.
37. Gatechompol S, Avihingsanon A, Apornpong T, et al. Efficacy and improvement of lipid profile after switching to rilpivirine in resource limited setting: real life clinical practice. AIDS Res Ther 2019;**16**(1):7.
38. Han WM, Apornpong T, Kerr SJ, et al. CD4/CD8 ratio normalization rates and low ratio as prognostic marker for non-AIDS defining events among long-term virologically suppressed people living with HIV. AIDS Res Ther 2018;**15**(1):13.
39. Han WM, Ubolyam S, Apornpong T, et al. Characteristics of suboptimal immune response after initiating antiretroviral therapy among people living with HIV with a pre-treatment CD4 T cell count <200 cells/mm(3) in Thailand. J Virus Erad 2020;**6**(3):100005.
40. Hiransuthikul A, Chutinet A, Sophonphan J, et al. Incidence and risk factors of ischemic stroke and transient ischemic attack among people living with HIV: a longitudinal cohort study. AIDS Res Hum Retroviruses 2021.
41. Putcharoen O, Wattanachanya L, Sophonphan J, et al. New-onset diabetes in HIV-treated adults: predictors, long-term renal and cardiovascular outcomes. AIDS (London, England) 2017;**31**(11):1535-43.
42. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet (London, England) 2004;**363**(9403):157-63.
43. Jin X, Han J. K-Means Clustering. In: Sammut C, Webb GI, eds. Encyclopedia of Machine Learning and Data Mining. Boston, MA: Springer US, 2017:695-97.
44. Yuan C, Yang H. Research on K-Value Selection Method of K-Means Clustering Algorithm. J 2019;**2**(2):226-35.
45. Kaoungku N, Suksut K, Chanklan R, et al. The Silhouette Width Criterion for

Clustering and Association Mining to Select Image Features. International Journal of Machine Learning and Computing 2018;**8**:69-73.

46. Wang X, Xu Y. An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. IOP Conference Series: Materials Science and Engineering 2019;**569**:052024.
47. A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters; 2006.
48. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;**19**(1):64.
49. Stamate D, Smith R, Tsygancov R, et al. Applying Deep Learning to Predicting Dementia and Mild Cognitive Impairment. Artificial Intelligence Applications and Innovations 2020;**584**:308-19.
50. Nadeau SE. Language and Aphasias. Continuum (Minneap Minn) 2021;**27**(6):1549-61.
51. Fridriksson J, den Ouden DB, Hillis AE, et al. Anatomy of aphasia revisited. Brain 2018;**141**(3):848-62.
52. Baldo JV, Katseff S, Dronkers NF. Brain Regions Underlying Repetition and Auditory-Verbal Short-term Memory Deficits in Aphasia: Evidence from Voxel-based Lesion Symptom Mapping. Aphasiology 2012;**26**(3-4):338-54.
53. Wicklund M. Clinical Approach to Cognitive and Neurobehavioral Symptoms. Continuum (Minneap Minn) 2021;**27**(6):1518-48.
54. Sacktor NC, Wong M, Nakasujja N, et al. The International HIV Dementia Scale: a new rapid screening test for HIV dementia. Aids 2005;**19**(13):1367-74.
55. Hinkin CH, Hardy DJ, Mason KI, et al. Medication adherence in HIV-infected adults: effect of patient age, cognitive status, and substance abuse. Aids 2004;**18 Suppl 1**(Suppl 1):S19-25.
56. Waldrop D, Irwin C, Nicholson WC, et al. The Intersection of Cognitive Ability and HIV: A Review of the State of the Nursing Science. J Assoc Nurses AIDS Care 2021;**32**(3):306-21.
57. Obermeit LC, Morgan EE, Casaletto KB, et al. Antiretroviral Non-Adherence is Associated With a Retrieval Profile of Deficits in Verbal Episodic Memory. Clin Neuropsychol 2015;**29**(2):197-213.

58. Heaton RK, Marcotte TD, Mindt MR, et al. The impact of HIV-associated neuropsychological impairment on everyday functioning. J Int Neuropsychol Soc 2004;**10**(3):317-31.

59. Patton DE, Woods SP, Franklin D, Jr., et al. Relationship of Medication Management Test-Revised (MMT-R) performance to neuropsychological functioning and antiretroviral adherence in adults with HIV. AIDS Behav 2012;**16**(8):2286-96.

VITA

NAME	Akarin Hiransuthikul
DATE OF BIRTH	14 March 1990
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Department of Preventive and Social Medicine, Faculty of Medicine, Chulalongkorn University
HOME ADDRESS	5 Soi 5 Seri 4 (Rama IX 43) Rd. Pattanakarn, Suanluang Bangkok, Thailand 10250
PUBLICATION	Hiransuthikul A, Chutinet A, Sophonphan J, Ubolyam S, Ruxrungtham K, Avihingsanon A. Incidence and risk factors of ischemic stroke and transient ischemic attack among HIV-infected individuals: a longitudinal cohort study. AIDS Res Hum Retroviruses. 2022;38(2):131-4. Hiransuthikul A, Janamnuaysook R, Getwongsa P, Peelay J, Samitpol K, Amatsombat T, Chumnanwet P, Chancham A, Kongkapan J, Rueannak J, Himma L, Srimanus P, Teeratakulpisarn N, Avery M, Wansom T, Mills S, Ramautarsing RA, Phanuphak N. Estradiol and testosterone concentrations among Thai transgender women in a transgender-led, integrated gender-affirming care and sexual health clinic: a real-world analysis. Transgend Health. 2021. Hiransuthikul A, Janamnuaysook R, Himma L, Taya C, Amatsombat T, Chumnanwet P, Samitpol K, Chancham A, Kongkapan J, Rueannak J, Getwongsa P, Srimanus P, Teeratakulpisarn N, Thammajaruk N, Avery M, Wansom T, Mills S, Ramautarsing RA, Phanuphak N. Acceptability and

Satisfaction towards Self-collection for Chlamydia and Gonorrhoea Testing among Transgender Women in Tangerine Clinic, Thailand: Shifting towards the New Normal. J Int AIDS Soc. 2021 Sep;24(9):e25801.

Hiransuthikul A, Chutinet A, Suwanwela NC. Ischemic stroke subtypes among Thai HIV-infected patients: a 12-year retrospective study. AIDS Res Hum Retroviruses. 2021;37(9):627-30.

Hiransuthikul A, Himma L, Kerr SJ, Janamnuaysook R, Dalodom T, Phanjaroen K, Pankam T, Kongkapan J, Mills S, Vannakit R, Phanuphak P, Phanuphak N. Drug-drug interactions among Thai HIV-positive transgender women undergoing feminizing hormone therapy and antiretroviral therapy: the iFACT study. Clin Infect Dis. 2021 Feb 1; 72(3): 396-402.

Hiransuthikul A, Chutinet A, Sakulrak S, Samajarn J, Vongsayan P, Kijpaisalratana N, Akarathanawat W, Apornpong T, Sangarlangkarn A, Gatechompol S, Han WM, Chattranukulchai P, Kerr S, Ruxrungtham K, Avihingsanon A. Carotid Intima-Media Thickness Is Not Associated with Neurocognitive Impairment Among People Older than 50 Years With and Without HIV Infection from Thailand. AIDS Res Hum Retroviruses. 2019 Nov/Dec; 35(11-12):1170-1173.

Siwamogsatham S, Chutinet A, Vongsayan P, Samajarn J, Putcharoen O, Aponpong T, Hiransuthikul A, Gatechompol S, Han WM, Chattranukulchai P, Kerr SJ, Ruxrungtham K, Avihingsanon A; HIV-NAT 227/HIV-NAT006 study team. Low CD4 Cell Counts Are Associated with Carotid Plaque and Intima-Media Thickness in Virologically Suppressed HIV-Infected Asians Older Than 50 Years. AIDS Res Hum Retroviruses. 2019 Nov/Dec;35(11-12):1160-1169.

Hiransuthikul A, Janamnuaysook R, Himma K, Kerr SJ, Thammajaruk N, Pankam T, Phanjaroen K, Mills S, Vannakit R, Phanuphak P, Phanuphak N. Drug-drug interactions between feminizing hormone therapy and pre-exposure prophylaxis among transgender women: the iFACT study. J Int AIDS Soc. 2019. 22(7):e25338.

Hiransuthikul A, Sungsing T, Jantarapakde J, Trachunthong D, Mills S, Vannakit R, Phanuphak P, Phanuphak N. Correlations of chlamydia and gonorrhoea among pharyngeal, rectal and urethral sites among Thai men who have sex with men: multicentre community-led test and treat cohort in Thailand. BMJ Open. 2019; 9(6): e028162.

Hiransuthikul A, Janamnuaysook R, Sungsing T, Jantarapakde J, Trachunthong D, Mills S, Vannakit R, Phanuphak P, Phanuphak N. High burden of chlamydia and gonorrhoea in pharyngeal, rectal and urethral sites among Thai transgender women: implications for anatomical site selection for the screening of STI. Sex Transm Infect. 2019 Nov; 95(7): 534-9.

Putcharoen O, Pleumkanitkul S, Chutinet A, Vongsayan P, Samajarn J, Sophonphan J, Kerr SJ, Hiransuthikul A, Siwamogsatham S, Ruxrungtham K, Avihingsanon A. Comparable carotid intima-media thickness among long-term virologically suppressed individuals with HIV and those without HIV in Thailand. J Virus Erad. 2019 Jan 1;5(1):23-27.

Hiransuthikul A, Trachunthong D, Pattanachaiwit S, Teeratakulpisarn N, Chamnan P, Pathipvanich P, Thongpaen S, Pengnonyang S, Nonenoy S, Lertpiriyasuwat C, Phanuphak P, Phanuphak N on behalf of the Thai MSM/TGW Test and Treat Study Group. Changes in risk

behaviors among Thai men who have sex with men and transgender women enrolled in the Test and Treat cohort. AIDS Care. 2019; 31(9): 1178-83.

Hiransuthikul A, Pattanachaiwit S, Teeratakulpisarn N, Chamnan P, Pathipvanich P, Thongpaen S, Pengnonyang S, Trachunthong D, Nonenoy S, Lertpiriyasuwat C, Phanuphak P, Phanuphak N on behalf of the Thai MSM/TGW Test and Treat Study Group. High subsequent and recurrent STIs prevalence among newly diagnosed HIV-positive Thai MSM and TGW in the Test and Treat cohort. Int J STD AIDS. 2019; 30(2): 140-6.

Hiransuthikul A, Thammajaruk N, Techatanawat I, Karachot B, Chuasuwan B, Manamuti C, Duereh M, Sawpitiporn M, Sophonphan J, Gatechompol S, Avihingsanon A, Bowonwattanuwong C, Ruxrungtham K, HIV-NAT 223 study team. Pharmacokinetics and 48-week safety and efficacy of generic ritonavir tablet-boosted atazanavir in HIV-1-infected Thai adults. Antivir Ther. 2018; 23(8): 699-703.

Han WM, Apornpong T, Kerr SJ, Hiransuthikul A, Gatechompol S, Do T, Ruxrungtham K, Avihingsanon A. CD4/CD8 ratio normalization rates and low ratio as prognostic marker for non-AIDS defining events among long-term virologically suppressed people living with HIV. AIDS Res Ther. 2018 Sep 27;15(1):13.

Hiransuthikul A, Wongkanya R, Sirivichayakul S, Trachunthong D, Sungsing T, Pankam T, Phanuphak P, Phanuphak N. Discordance in drug-resistance mutations between blood plasma and semen or rectal secretion among newly diagnosed HIV-1-infected Thai MSM. AIDS Res Hum Retroviruses. 2018; 34(7): 626-8.

Hiransuthikul A, Hiransuthikul N. Salmonella septic bursitis

at ankle region – an uncommon feature of septic bursitis in an HIV-infected patient: a case report and literature review. Southeast Asian J Trop Med Public Health. 2016; 47(6): 1177-82.

Hiransuthikul A, Rattananupong T, Klaewsongkram J, Rerknimitr P, Pongprutthipan M, Ruxrungtham K. Drug-induced hypersensitivity syndrome/drug reaction with eosinophilia and systemic symptoms (DIHS/DRESS): 11 years retrospective study in Thailand. Allergol Int. 2016; 65(4): 432-8.

Hiransuthikul A, Hiransuthikul N. Cytomegalovirus gastrointestinal disease in patients with AIDS: A case series. Chula Med J 2016 Mar - Apr; 60(2): 155-65.

AWARD RECEIVED

2022 International Scholarship Award, American Academy of Neurology

Best resident research, Graduate Affairs, Faculty of Medicine, Chulalongkorn University

Best neurology resident/fellow research, The Neurological Society of Thailand

IAS 2021 Scholarship, 11th IAS Conference on HIV Science 2021

3rd-place best abstract award, Asia Pacific Stroke Conference 2020

Speaker at AIDS 2018 Official Press Conference on Prevention Research Highlights

New Investigator Award, 2018 Conference on Retroviruses and Opportunistic Infections

Young Investigator Award, 2017 Conference on Retroviruses and Opportunistic Infections

Conference Participation Grant, 1st Asia HIV Clinical Forum & 2nd APACC