February 2022

Dissertation for Ph.D.

A Research on Machine Learning based Paging Enhancement in 5G Network

Graduate School of Chosun University

Department of Information and Communication Engineering

Wan-Kyu Choi

A Research on
Machine Learning based
Paging Enhancement in 5G Network

5G 네트워크에서 머신 러닝을 적용한
페이징 향상에 관한 연구

February 25, 2022

Graduate School of Chosun University

Department of Information and Communication Engineering

Wan-Kyu Choi

A Research on Machine Learning based Paging Enhancement in 5G Network

Advisor: Prof. Jae-Young Pyun

This thesis is submitted to the Graduate School of Chosun University in partial fulfillment of the requirements for the Doctor's degree in engineering.

October 2021

Graduate School of Chosun University

Department of Information and Communication Engineering

Wan-Kyu Choi

This is to certify that the Doctor's thesis of

Wan-Kyu Choi

has been approved by the examining committee for the thesis requirement for the Doctor's degree in engineering

Committee Chairperson **Chosun University**	**Prof. Goo-Rak Kwon (Sign)**
Committee Member **Chosun University**	**Prof. Jae-Young Pyun (Sign)**
Committee Member **Chosun University**	**Prof. Ji-Hwan Moon (Sign)**
Committee Member **Sunchon National University**	**Prof. Eui-Sung Kang (Sign)**
Committee Member **Sunchon National University**	**Prof. Sang-Hyun Park (Sign)**

January 2022

Graduate School of Chosun University

Table of Contents

List of Tables

List of Figures

Abbreviations

3GPP	Third Generation Partnership Project
5GS	5G System
5GC	5G Core Network
CP	Control Plane
DP	Data Plane
EPC	Evolved Packet Core
eNodeB	Evolved NodeB
gNodeB	Next Generation NodeB
IoT	Internet of Things
KNN	k-Nearest Neighbors
LTE	Long-Term Evolution
MME	Mobility Management Entity
AMF	Access and Mobility Management Function
SMF	Session Management Function
NGAP	Next Generation Application Protocol
NR	New Radio
RAN	Radio Access Network
RRC	Radio Resource Control
TA	Tracking Area
TAI	TA Identification
UE	User Equipment
AR	Augmented Reality
VR	Virtual Reality

Abstract

A Research on Machine Learning based Paging Enhancement in 5G network

Wan-Kyu Choi
Advisor: Prof. Jae-Young Pyun, Ph.D.
Depart. of Info. and Comm. Eng.,
Graduate School of Chosun University

Recently, technologies such as big data, artificial intelligence, and machine learning have been applied to intelligently and effectively operate fourth-generation (4G) and fifth-generation (5G) network systems. In particular, attentions are obtained from using those technologies in 4G mobility management entity (MME) and 5G access and mobility management function (AMF), where functional enhancement or performance improvement is required. This thesis shows an enhanced paging approach based on supervised machine learning and a Markov process for the performance improvement of paging in 5G AMF. User equipment (UE) profile information in 5G AMF classifies subscribers into two types using a UE classifier model with k-nearest neighbors (KNN) supervised learning. In this thesis, UE movement data between next-generation NodeBs (gNodeB)s are collected and analyzed, and the Markov process is applied to construct a transition probabilistic model. When a UE moves from an adjacent gNodeB in 5G connection management (5G CM)-Idle state, a method for predicting the gNodeB location is required to perform paging effectively on the predicted gNodeBs. In the proposed paging method, the AMF applies the UE profile information to the KNN-supervised learning model and classifies the subscriber UE type. In addition, based on the UE movement data obtained from the gNodeBs, it generates a probabilistic gNodeB list and then performs paging on the gNodeB list that is optimally selected from the combination of the subscriber UE type and the probabilistic gNodeB list. Experimentally, the paging response and signals of the proposed method are shown in comparison with the existing paging method for finding a UE using the recently visited gNodeB list in the tracking area (TA) of the AMF. This thesis shows a new paging method with the probability procedure of Markov process and the classification of the kNN-supervised machine learning, analyzes and examines the results of the performance improvement accordingly.

요 약

5G 네트워크에서 머신 러닝 적용한 페이징 향상에 관한 연구

최 완 규
지도교수: 변 재 영
조선대학교 대학원 정보통신공학과

최근 4차 산업 혁명 시대에 적용되는 빅 데이터, 인공 지능, 머신 러닝 등의 기술은 fourth-generation (4G)와 fifth-generation (5G)의 코어 네트워크에서 상용 시스템들이 보다 지능적이고 효과적으로 시스템을 운영할 수 있도록 여러 방면으로 적용되고 있는데, 특히 4G mobility management entity (MME)와 5G access and mobility management function (AMF)에서도 자체적으로 기능 향상이나 성능 개선이 요구되는 분야에 활용하는 방안들이 적극적으로 논의되고 있다. 본 논문은 5G AMF에서 기존 페이징 방법에 대한 성능 개선 방법으로 지도 기계 학습과 마코프 프로세스를 기반으로 머신 러닝 페이징 방법을 적용하였다. 5G AMF에서 user equipment (UE) 프로파일 정보는 k-nearest neighbors (KNN) 지도 학습을 적용한 UE분류기 모델을 사용하여 가입자를 2가지 유형으로 분류하였고, 그리고 next generation NodeB (gNodeB)간의 UE 이동 데이터를 수집 및 분석 하고 마코프 프로세스를 적용하여 gNodeB별 전이 확률 모델을 구성하였고, UE가 5G connection

management (5G CM)-Idle 상태에서 인접 gNodeB로 이동하는 경우에, AMF는 가입자 유형에 따라서 UE가 존재할 가능성이 높은 gNodeB를 예측하고, 최종적으로 선택된 gNodeB에 페이징을 수행하는 방안을 제안하였다.

제안된 페이징 방법에서 AMF는 UE의 프로파일 정보를 KNN 지도 학습 모델에 적용하여 UE의 가입자 유형을 분류하고, 그리고, gNodeB들간의 UE 이동 데이터를 모두 수집 및 분석 하고, 이 결과를 기반으로 마코프 프로세스를 적용한 확률적 gNodeB 리스트를 생성하였고, 최종적으로 두 가지 방법의 결과를 조합하여 선택된 gNodeB 리스트에 페이징을 수행하였다. 실험을 통해서 제안된 페이징 방법은 AMF의 tracking area (TA)에서 최근에 방문한 gNodeB 리스트를 이용하여 UE의 위치를 찾는 기존의 페이징 방법과 비교하여 빠른 페이징 응답과 전체 페이징 시그널 횟수의 감소를 확인하였다. 본 논문에서는 지도 기계 학습의 분류와 마코프 프로세스의 확률 과정을 적용한 새로운 페이징 방법과 절차를 보여주고, 그에 따른 페이징 성능 개선의 결과에 대해서 분석 및 고찰 한다.

Ⅰ. Introduction

A. Research Background and Motivation

Technologies leading the recent Fourth Industrial Revolution, such as big data, artificial intelligence, and Internet of Things (IoT), are expected to play an important role in fifth-generation (5G) networks. By 2025, the number of 5G subscribers is expected to reach 26 billion people [1].

In particular, 5G requires ultra-high-speed, hyper-connectivity, and ultra-low latency to link all objectives and provide interworking among various services through the network. Therefore, data transmission via an ultra-wide bandwidth is mainly necessary, and the support of a large number of IoT devices is required to seamlessly exchange information simultaneously on the network [2]. Ultra-low latency is also a necessary technology that ensures self-driving by safely operating a car in an autonomous driving environment and responding to unexpected situations by minimizing the delay time for data transmission.

In 5G mobile communication, which includes more IoT devices compared with the existing fourth-generation 4G mobile communication, providing ultra-low latency connection to user equipment (UE) and the supporting numerous IoT devices are critical in both the core networks and radio access network (RAN). If the UE and IoT devices required to transmit data from the access and mobility management function (AMF) to the UE while maintaining the connection management (CM)-Idle and registration management-Registered states in both next-generation NodeB (gNodeB) and AMF, then the AMF notifies its transmission the UE through the gNodeB by performing a paging process before data transmission [3]. When the

UE moves to an adjacent gNodeB, it is registered with the corresponding gNodeB. However, when assuming that the UE moves frequently between gNodeBs, the mobility registration procedure of the AMF between gNodeB and the UE consumes a significant amount of unnecessary radio resource. Because not all mobility registration procedures of UE are reported to the AMF, the latter is expected to page multiple gNodeBs for the actual connection with its managed UE, which degrades the performance of the entire mobile network, e.g., increased signal processing load and increased latency. To solve these problems, some works on predicting the location of the UE and selecting optimal gNodeBs for paging in 4G evolved packet core (4G EPC)/ 5G core (5GC) networks have been actively conducted.

Some methods proposed in the related studies are as follows. By utilizing the location movement information of the user terminal, a TA was overlappingly allocated to several adjacent TA lists, then performed the paging a small TA list selectively and the paging load was reduced [4]. As an alternative to the method of performing paging directly on numerous eNodeBs in the existing mobility management entity (MME), the study has been also conducted that a method is performed that an eNodeB conducts paging to other eNodeBs for reducing the paging load of the MME [5]. Recently, 3rd generation partnership project (3GPP) has designated a node such as the network data analytics function (NWDAF) that performs various analyzing on mobility for the UE on a separated external node of the network to help in the selection of the optimal RAN in the actual paging process [6]. There is also a technique of bundling the base stations in the TA, it is divided the base stations into groups, and applied machine learning using subscriber data, and paged the group of the base station where the UE location is predicted [7]. Even when the results of previous research are applied to the actual 4G/5G commercial network system, it is important to confirm that they do not affect the operation of the commercial mobile communication system and have the paging signal reduction, but no effect of the increasing delay of paging response that expected in the previous their research.

B. Research Objective

In the studies pertaining to the paging procedure in the 4G network, various methods afford improvements that reduce the performance degradation of the MME owing to the paging process. The results of improvements such as load reduction and delay reduction, which were predicted in a related study, have not been proven through application to actual commercial 5G systems. Hence, the improvement in 5G networks was not confirmed.

Herein, the effect of applying machine learning to the AMF paging process in a 5G core network is discussed. Paging is the process of determining the location of the UE and is used to inform the UE of subscribers regarding the arrival of the voice or data to be delivered to the UE in the CM-idle state. In the 5G AMF-initiated paging procedure, when the AMF receives the N1N2MessageTransfer signal from the session management function (SMF), it transmits the next generation application protocol (NGAP) paging signal to the gNodeB in the TA. Subsequently, when the AMF does not receive a response from the gNodeB to the completion of the connection of radio resource with the UE within a certain duration, the AMF retransmits the NGAP paging signal to the gNodeBs in all TAs. The maximum number of paging attempts was set in the paging profile configuration of the AMF, and the proportion of these paging signals constituted approximately 30% of the total signals processed in the AMF, which is an important target of performance improvement for the AMF and imposes a significant burden on the load as well as the linked gNodeB simultaneously. Therefore, in a 5G wireless communication environment that requires hyper-connectivity and ultra-low latency with the UE, the paging process must be improved by reducing the paging signal load of the AMF and gNodeB.

A probabilistic paging method using supervised machine learning and a Markov process instead of the typical paging method is proposed and introduced herein. First, In Chapter 2 describes the background knowledges and the existing paging procedure and Chapter 3 presents the proposed paging method and procedure using supervised machine learning and the Markov process. Chapter 4 describes the experimental environment and provides an evaluation of the proposed method based on experimental results. Finally, Chapter 5 presents the conclusions of the proposed method.

Ⅱ. Background and Related works

A. 5G Network

5G refers to the 5th generation mobile communication standard that achieves digital transformation that will lead the IoT era beyond the prediction that the explosively increasing amount of data will be difficult to handle with 4G technology in the future. According to the definition given by the International Telecommunication Union (ITU), 5G is a mobile communication technology with a maximum download speed of 20 Gbps and a minimum download speed of 100 Mbps and is a key infrastructure of the Fourth Industrial Revolution [8]. The exact name of 5G is IMT-2020, which is about 20 times faster and has 100 times more processing capacity compared with 4G currently used in daily life, thus it is expected that the actual performance will be improved by more than 10 times.

Table 1. Comparison of 4G and 5G technologies [9]

Compared item	4G	5G
Maximum transmission speed	1 Gbps	20 Gbps
User experience transmis-sion speed	10 Mbps	100–1,000 Mbps
Allowable maximum mobility speed	350 km/h	500 km/h
Latency	10 ms	1 ms
Maximum connecting in-strument	100,000 /km^2	1,000,000 /km^2
Data processing capa. per the area	0.1 Mbps/m^2	10 Mbps/m^2
Power efficiency	1×	8,100×

Although consumers may have low expectations for 5G since a lot of people do not have any inconvenience of living only in a 4G-based smart communication environment, but the coming 5G era has a lot of potential for development in artificial intelligence (AI), big data, augmented reality/virtual reality (AR/VR) and Cloud, thus it is expected that a significantly different world will come [10]. Table 1 shows the differences between 4G and 5G networks by comparing various items [11].

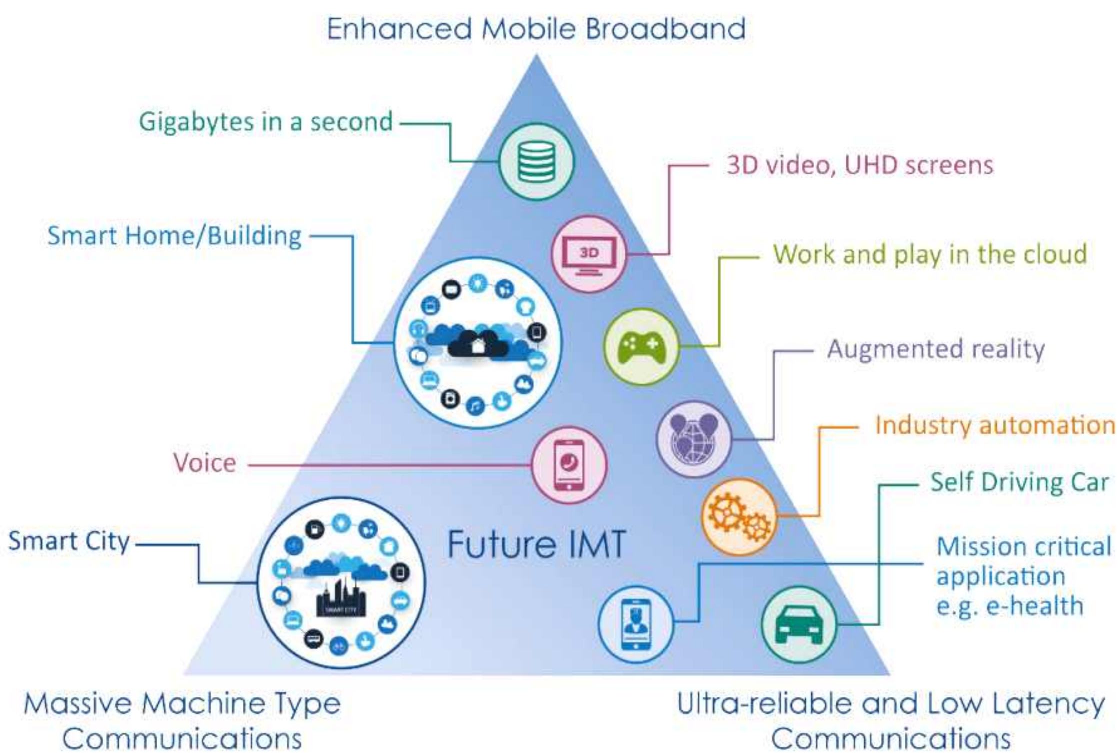


Figure 1. 5G Usages and Requirements [8]

As shown in Figure 1, 5G mobile communication technology is targeting three major technological evolution directions [12].

(1) Ultra-High Data Rate(eMBB: enhanced Mobile Broadband)

(2) Fast and Highly Reliable(URLLC: Ultra Reliable & Low Latency Communications)

(3) Massive Connection(mMTC: massive Machine-Type Communications)

First, eMBB enables the implementation of 4K, 8K, hologram, AR/VR, etc. with no difference between wired and wireless in 5G, enabling a communication environment that can use high-capacity and high-speed data [13]. That is, immersive contents such as 4K,8K, AR/VR are implemented in multimedia devices and high-resolution and full-screen TVs do not need to be fixed in a specific place due to a wired set-top box. When downloading a ultra high definition (UHD) movie of 15 giga byte (GB) size, it takes 240 seconds in the latest 4G at 500 Mbps, while it takes 6 seconds in 5G at 20 Gbps [14]. In particular, it aims to provide a speed of 100Mbps not only in areas with strong signals near base stations but also in areas with weak signals. In this way, a UHD streaming service without interruption will be possible even in crowded places where users are dense, such as downtown areas where thousands of people come and go or stadiums where major sports are held [15].

Second, mMTC defines as increasing the number of IoT devices by a maximum of 1 million per unit area and reducing the price of the device, extending long battery life by 10 years, and expanding service coverage [16]. The maximum number of connecting devices in 5G is more than 10 times long term evolution (LTE) and various devices and sensors can be used to provide a variety of services. As a high-density device connection environment, a sensor network that is applied to meter reading, agriculture, buildings, and logistics is implemented, and finally, a real IoT is possible to use. IoT can also be divided into two types based on its service characteristics. First, the multi-device connectivity IoT (Massive IoT), which allows connection to many devices at a low price to collect various information and control the devices, second, the ultra-stable IoT (Mission Critical IoT) that can be used in medical and self-driving vehicles [17]. In the past, the dedicated IoT networks of 4G focused on building an environment where many devices can be connected at a low price, 5G also enables the ultra-stable IoT service. 5G is expected to enable services to new IoT areas that were not included in the existing 4G IoT dedicated networks [18].

Third, uRLLC means that the latency time for the data sent from the smartphone to return to the UE via the base station, mobile switch center, and server is very short. When moving from 4G to 5G, the transmission latency is reduced by about one-tenth from 10ms of LTE to 1ms. With this help, the ‘real-time’ service that was previously impossible in 4G becomes possible [19]. For example, when a self-driving vehicle is autonomously driving at a speed of 100 Km/h, if a command occurs to emergency breaking from a server, then it takes times for, assuming a 50 ms delay in 4G, a self-driving vehicle receives a stop signal after a 1.4 m delay, while assuming a 1 ms delay in 5G, receives a stop signal after a 2.8 cm delay.

In the 5G environment, the network slicing technology enables the above mobile broadband and IoT to be simultaneously implemented with a single technology and network and provides various services, as shown in Figure 2 [20, 21].

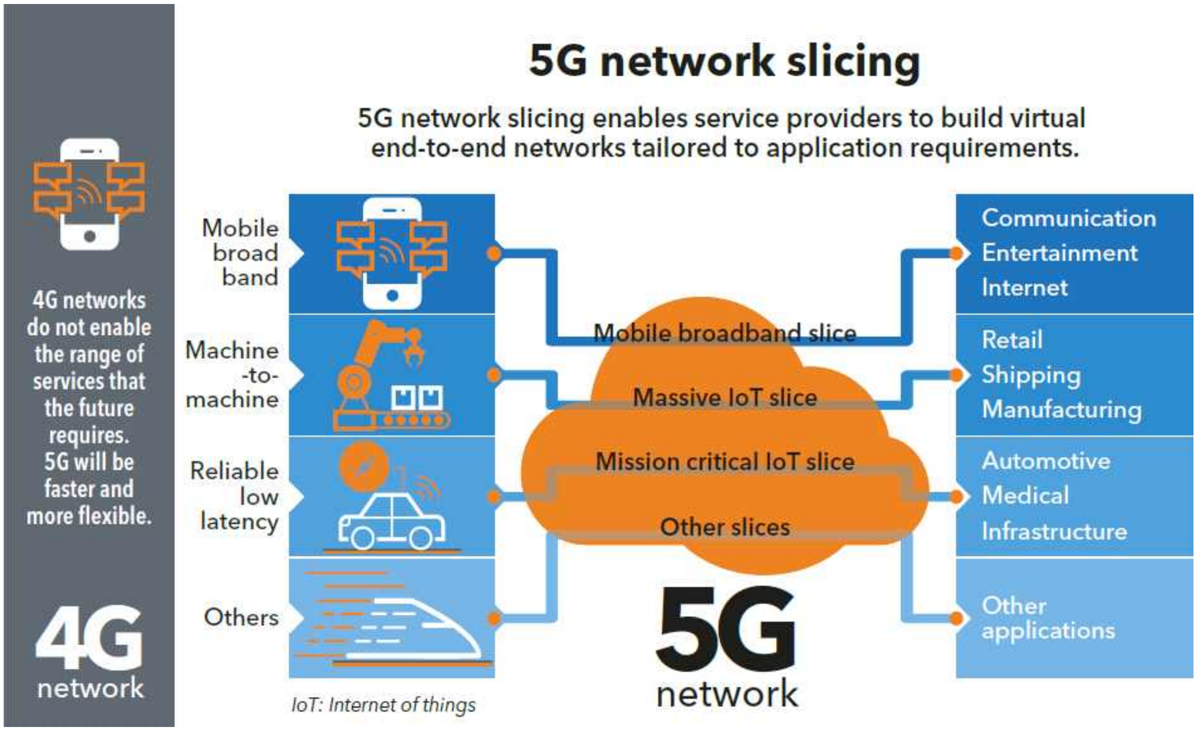


Figure 2. 5G Network Slicing [22]

If these changes are made, the IoT dedicated networks that could not be included in 4G will not need to exist in the 5G era. In addition, services using mobile broadband networks such as smartphones, services using wired broadband networks such as IPTV and IoT services can be managed in one, making it easier to implement self-driving vehicles, factory automation, virtual and augmented reality, and tele-medicine. In the IMT-2020 standard, a delay time of less than 1 millisecond (ms) and a data transmission packet error rate of 10^{-5} is set as technical performance requirements for URLLC.

In the 5G standard, both non-standalone (NSA) and standalone (SA) structures for 5G deployment are considered for the core network for evolution from 4G to 5G [23].

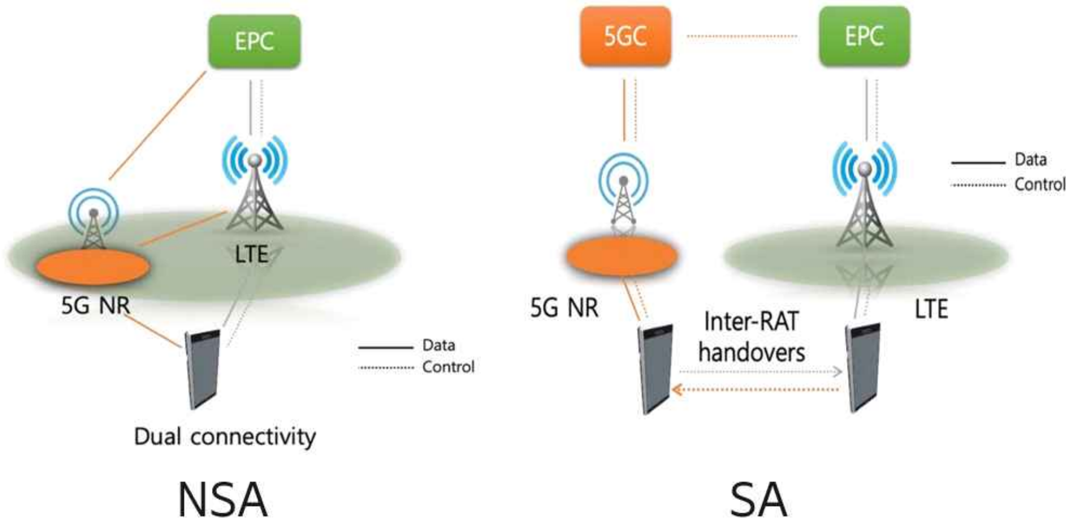


Figure 3. 5G NSA and SA Architecture [24]

In Figure 3, NSA was implemented in the initial 5G commercial network and most 5G commercial services are currently provided in the NSA structure. The control plane (CP) function to manage the mobility of the 5G UE utilizes 4G LTE network and the 5G new radio (NR) which handles the data plane (DP) function likes as the traffic of the 5G UE transmits data through the 4G EPC. It has the advantage of efficiently using the existing 4G LTE and realizing the 5G service and the disadvantage of being concerned about overloading the EPC with a complex network structure through EPC inter-working. The SA mode is a structure using the 5G NR network and the 5GC structure for both the control channel and the data channel. All 5G services such as self-driving, ultra-low latency, VR/AR, and hyper-connectivity are available and a nationwide 5G network is required. In both structures, the UE evolves to support both 4G and 5G wireless access at the same time [25].

5G Core Network

The 5G core network, which enables the advanced functionality of 5G networks, is one of three primary components of the 5G system, also known as 5G System (5GS). The other two components are the 5G access network (5G-AN) and user equipment (UE). The 5G core (5GC) uses a cloud-aligned service-based architecture (SBA) to support authentication, security, session management, and aggregation of traffic from connected devices, all of which requires the complex interconnection of network functions (NF) and each NF offer one or more services to other NFs via service-based interface (SBI), as shown in Figure 4.

The components of the 5G core architecture normally include user plane function (UPF), data network (DN), access and mobility management function (AMF), authentication server function (AUSF), session management function (SMF), network slice selection function (NSSF), network exposure function (NEF), NF repository function (NRF), policy control function (PCF), unified data management (UDM), and application function (AF).

5G network functions are split up by service. 5GS architecture is a service-based representation in which the control plane network functions access each other's services and is a reference point representation in which the interaction between the network functions is shown with point-to-point reference points. Therefore, 5G architecture is also defined as a service-based architecture.

The 5G network topology diagram in Figure 4 shows the major components and service-based interfaces of the 5G core network.

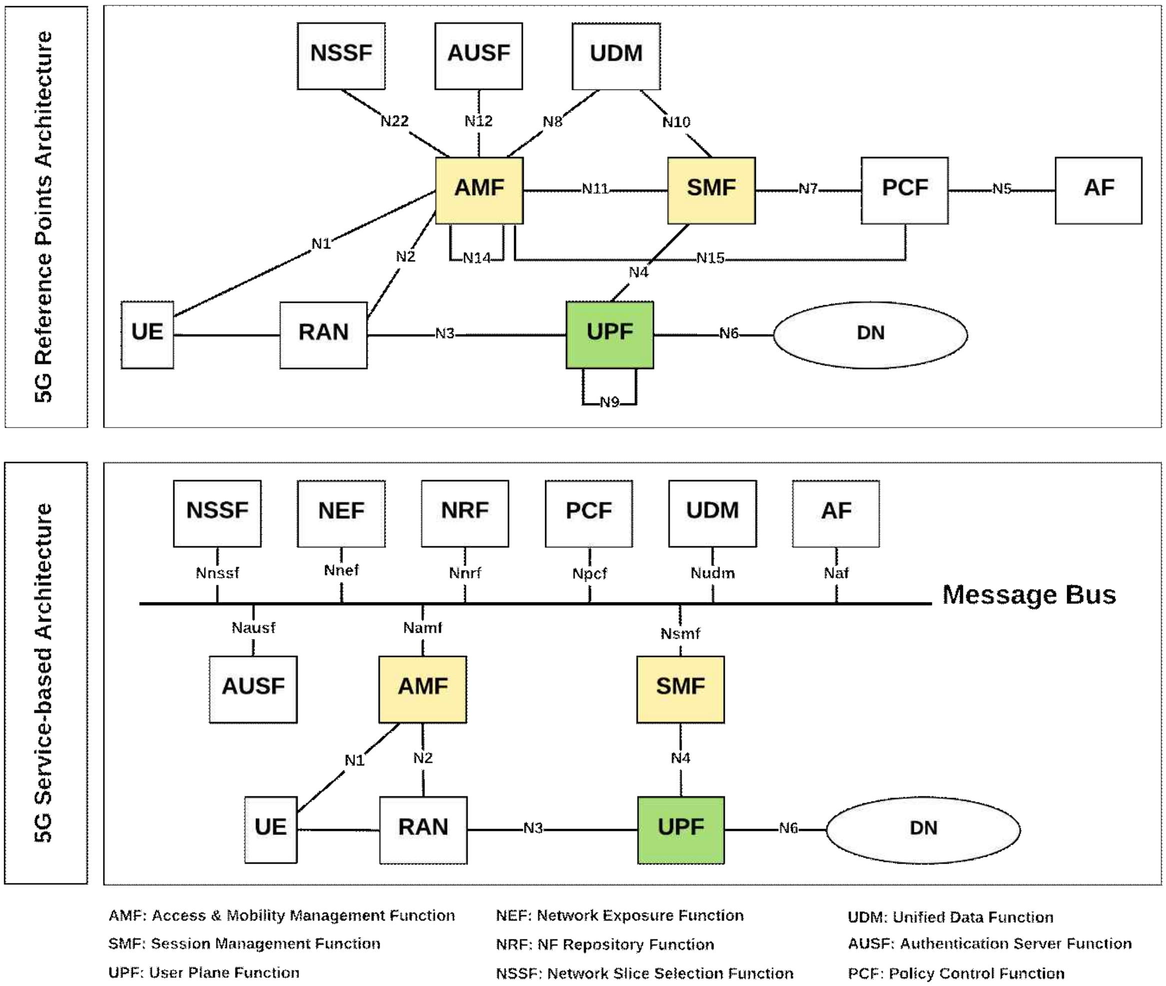


Figure 4. 5G System Architectures [26]

UE (User Equipment) like 5G smartphones or 5G cellular devices connect over the 5G New Radio Access Network to the 5G core and further to Data Networks, like the Internet.

gNodeB (Next Generation Node Base-station) is a 3GPP-compliant implementation of the 5G-NR base station. 5G wireless base stations transmit and receive communications between the user equipment and the mobile network.

AMF (Access and Mobility Management Function) acts as a single-entry point for the UE connection and is responsible for following. Termination of RAN Control Plane interface, Termination of NAS, NAS ciphering and integrity protection, Mobility Management, Lawful Intercept, Transparent proxy for routing access authentication and SM messages, Access Authentication, Access Authorization.

UPF (User Plane Function) transports the IP traffic data between the UE and the external networks and supports it as follows. Functions are QoS handling for User plane, Packet routing & forwarding, Packet inspection and Policy rule enforcement, Lawful Intercept, and Traffic accounting and reporting.

SMF (Session Management Function) supports the session management, UE IP address allocation & management, the selection and control of User Plane function, the termination of interfaces towards Policy control and Charging functions, the control part of policy enforcement and QoS, the lawful intercept, the termination of Session Management parts of NAS messages, the Downlink Data Notification, the roaming functionality, and the charging data collection and charging interface.

DN (Data Network): Operator services, Internet access, or other services.

AUSF (Authentication Server Function) allows the AMF to authenticate the UE and access the services of the 5G core, stores authentication keys and facilitate security processes.

UDM (Unified Data Management) stores the long-term security credentials used in authentication for AKA and the subscription information.

PCF (Policy Control Function) supports of unified policy framework to govern network behavior and policy rules to control plane function that enforces them.

NRF (Network Repository Function) supports a key element of the new 5G Service Based Architecture and provides service discovery between individual network functions such as the UPF, AMF, SMF, PCF, etc.

NSSF (Network slice selection function) selects the set of network slice instances serving the user equipment and determines which access and mobility management function to use.

B. 5G Paging

In 5G system, the number of gNodeBs is greater than that in 4G LTE owing to 5G service characteristics such as hyper-connectivity and 5G new radio (NR) frequency band features. As show in Figure 5, the AMF manages the gNodeB, TA, and TA list by consolidating multiple TAs, which comprise several adjacent gNodeBs. The UE initially sends a Registration Request to the AMF and then receives a success message of Registration and the TA list managed by the AMF. When the UE moves to a new TA that is not in the TA list, it performs a Registration procedure with its AMF and receives a new TA list [27].

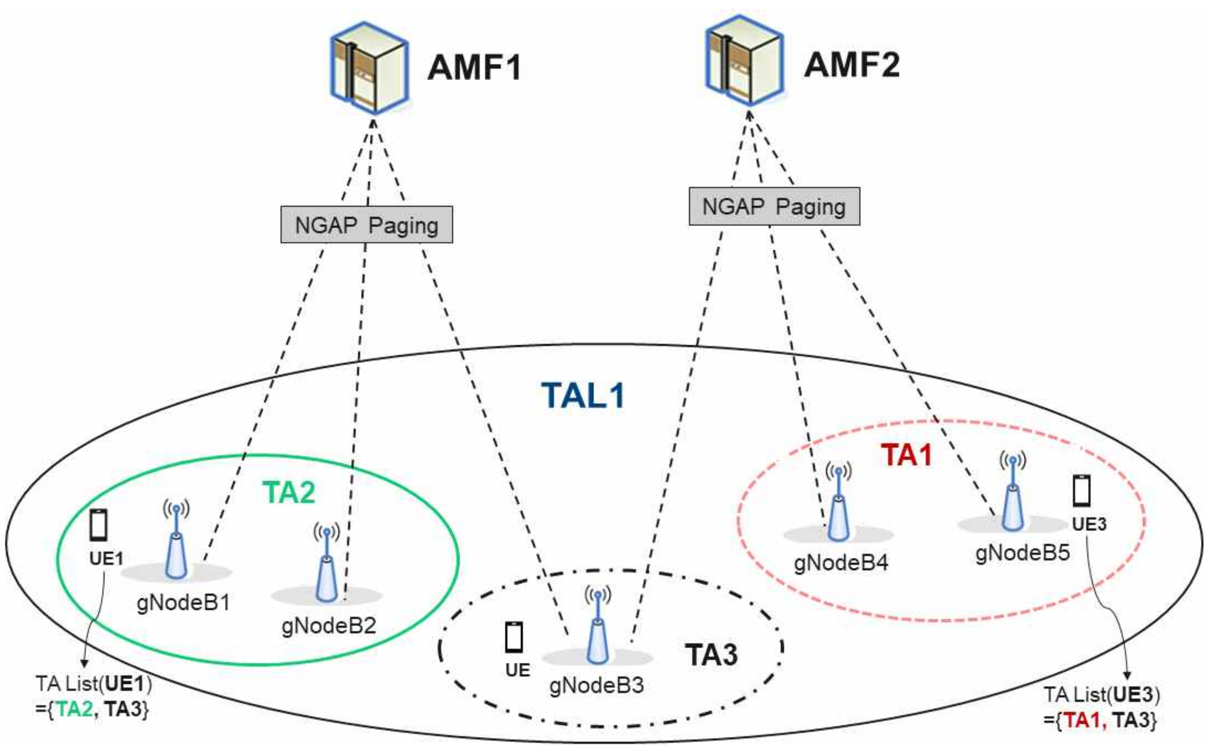


Figure 5. 5G NGAP Paging

In general, when the UE is in the CM-connected state, the AMF knows the location information, such as the gNodeB and cell information of the UE. If traffic data are transmitted to the UE, they may be directly transmitted to the UE via the connected channel of the radio resource. However, when the UE is in the CM-idle state, the AMF contains only the UE's last visited gNodeB, latest visited gNodeBs, and TA [28]. When the UE moves among the TAs, the UE reports its TA change to the AMF via the Mobility Registration Update Procedure, which is periodically operated based on the Periodic Registration Update Timer received from the AMF in the Registration Procedure. Subsequently, the AMF stores a new TA in the last visited gNodeBs of the UE and updates the TA list [29]. Next, the AMF can notify the UE of the data transmission or voice call when it is ready for transmission to the UE in the CM-idle state. As shown the protocol stack structure between the AMF, gNodeB, and UE in Figure 6, the paging messages are transmitted to the gNodeB based on NGAP and the gNodeB relay to the radio resource protocol and transmit to the UE [3].

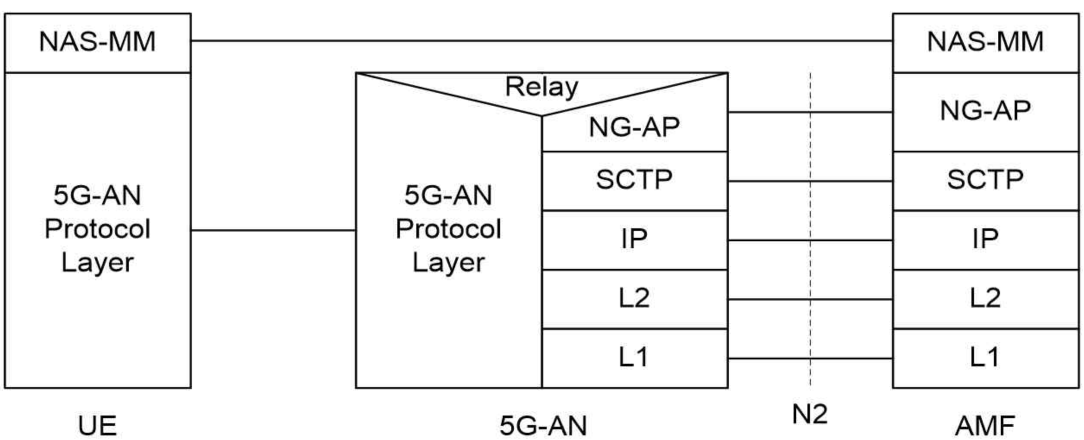


Figure 6. Protocol stack between UE and AMF

Legend:

- NAS-MM: The NAS protocol for MM functionality supports registration management functionality, connection management functionality, and user plane connection activation and deactivation. It is also responsible for ciphering and integrity protection of NAS signaling [27].

- 5G-AN Protocol layer: This set of protocols/layers depends on the 5G-AN. In the case of NG-RAN, the radio protocol between the UE and the NG-RAN node (gNodeB) is specified in Technical Specification [30,31].

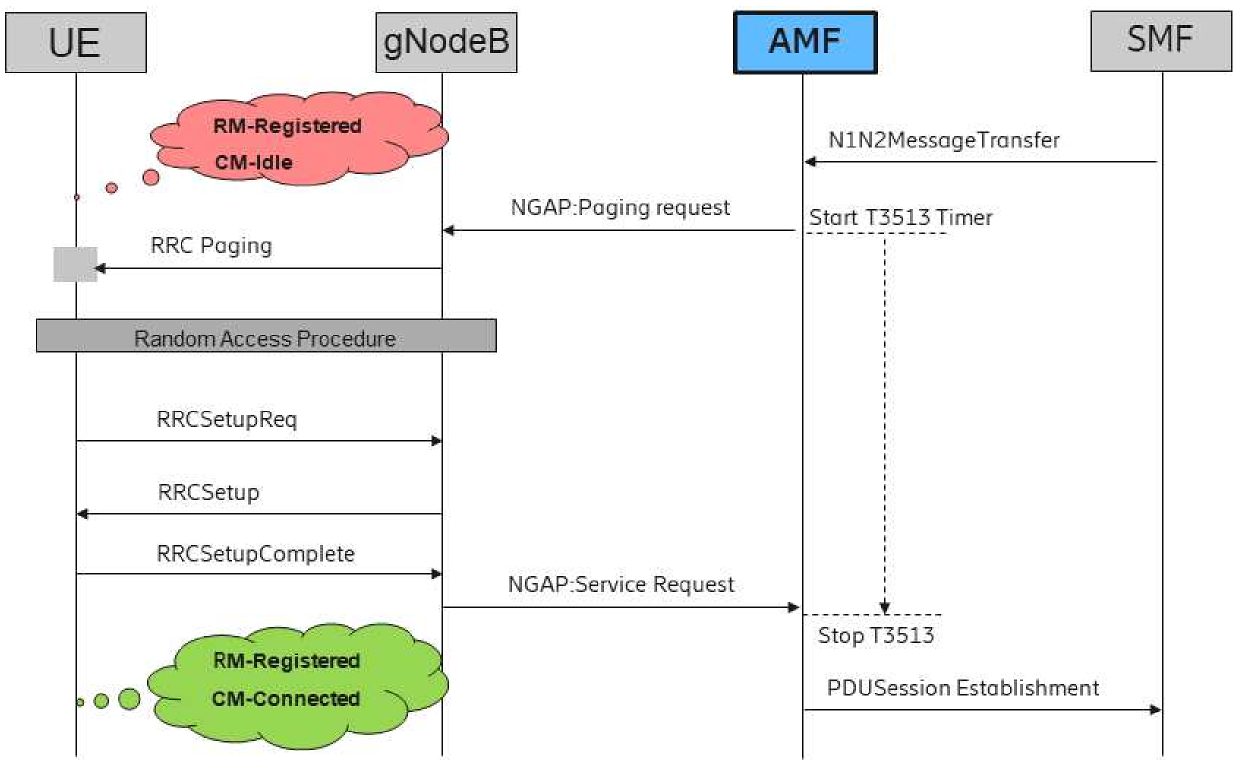


Figure 7. Sequence Flow of 5G AMF-Initiated Paging

The AMF paging process in Figure 7 shows the entire paging flow in the UE, gNodeB, AMF, and SMF [32]. As shown in Figure 7, the AMF initiates paging when it receives an N1N2MessageTransfer message from SMF. The AMF verifies the connection status of the corresponding UE. At this time, if the UE is in the CM-idle state, the AMF selects gNodeBs in which the UE is likely to exist, and transmits paging signals based on the NGAP to the gNodeBs. Subsequently, these gNodeBs broadcast paging signals to all cells via a paging channel. The UE monitors the wireless paging channel based on the gNodeB. If paging occurs that identifies the UE, then the state is changed to the CM-connected state with the gNodeB; subsequently, the traffic data are transmitted from the user plane function (UPF) to the UE [33].

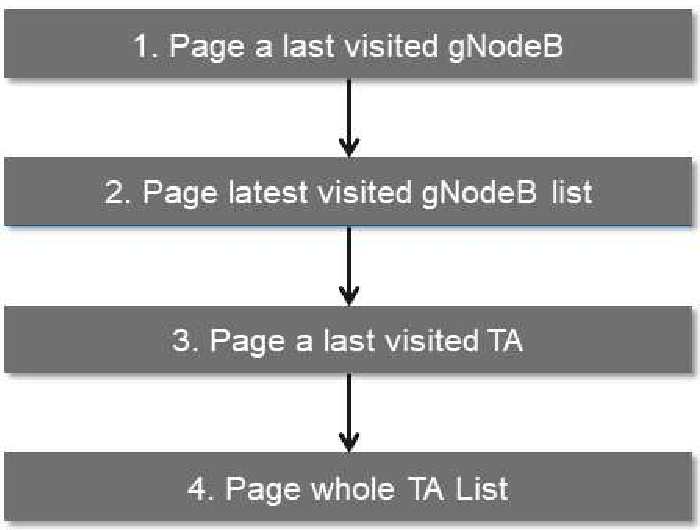


Figure 8. Typical AMF Paging Range Profile

To perform general paging in AMF, the latter manages the registered gNodeB, TA, and TA list; the paging profile configuration is shown in Table 2. The paging profile configuration has various paging-related settings for different UE service types. The configuration includes the paging range, number of paging attempts, and a paging profile created by the UE service type.

Table 2. Example of AMF Paging Profile Configuration

Paging Profile	Last Visited gNodeB	Latest Visited gNodeB List	Last Visited TA	TA List
1	0	0	0	4
2	0	0	2	3
3	0	1	2	2
...	...	...	...	...
20	3	2	2	2

Figure 8 shows an example of the paging range profile when No. 20 of the paging profile configuration in Table 2 is selected. The first paging message based the configuration is sent up to three times to the last visited gNodeB. If no response is received from the gNodeB within a certain duration, then the same paging message is sent up to two times to the gNodeBs in the latest visited gNodeB list as the second attempt. Subsequently, the third paging attempt can be sent up to two times to all the gNodeBs in the last visited TA when no response is obtained from the gNodeBs.

As the last paging attempt, a paging message is sent up to two times to the gNodeBs of all TA list. In other words, the number of paging attempts and the range of gNodeBs for the paging are set in the AMF paging profile configuration [34].

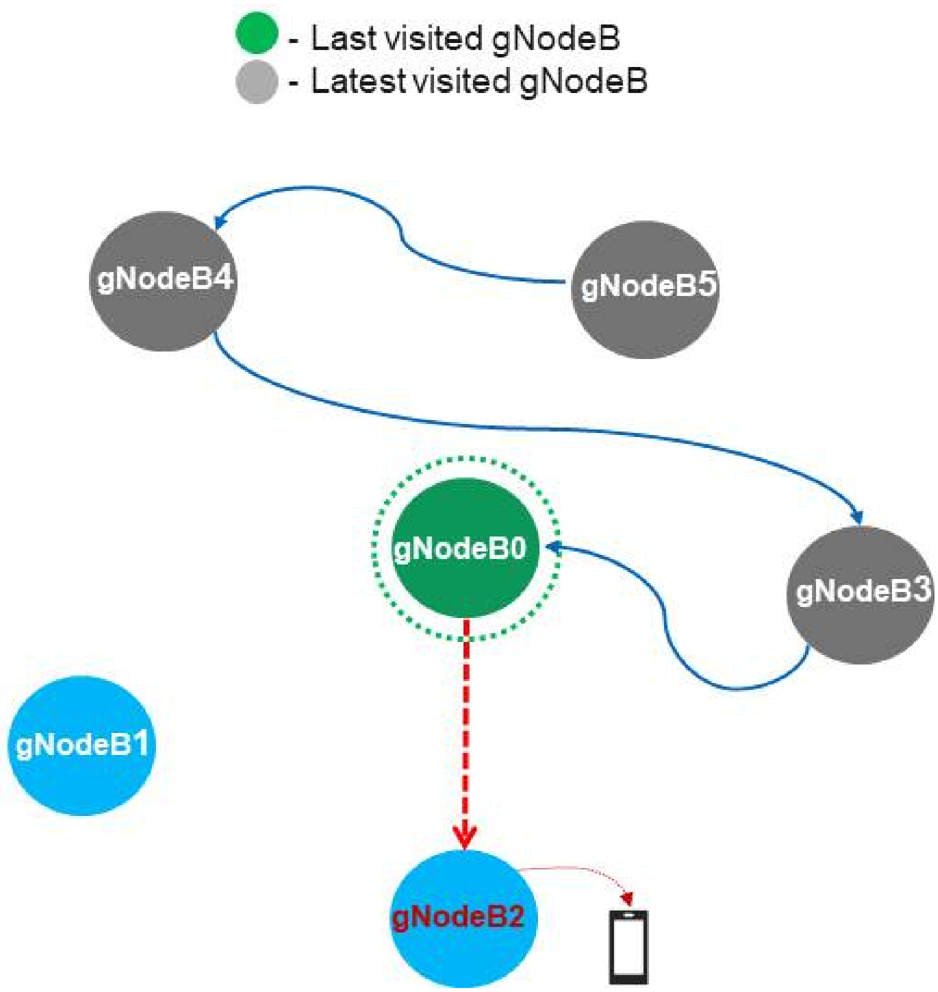


Figure 9. Example of paging on latest gNodeB list

Figure 9 shows an example of UE movement between gNodeBs within the same TA and the typical paging range in an AMF. Based on the paging range in Figure 8 and paging profile configuration No. 20 in Table 2, the first paging message is sent to gNodeB0, the second to gNodeB0, gNodeB3, gNodeB4, and gNodeB5, and the third to all gNodeBs from gNodeB0 to gNodeB5 in the TA. Finally, the UE responds during the third paging attempt.

C. Machine Learning

Machine learning was defined by Arthur Samuel in 1959 as “a field of study that develops algorithms that enable machines to learn from data and execute actions that are not individually specified in code” [35]. Machine learning is the study of computer algorithms that automatically improve through experience, a field of artificial intelligence, and the field of developing algorithms and technologies that enable computers to learn [36]. Machine learning is a technology that studies and builds algorithms and systems that learn based on empirical data, perform predictions, and improve their performance. Machine learning algorithms choose to build specific models to derive predictions or decisions based on input data, rather than performing strictly fixed static program instructions [37].

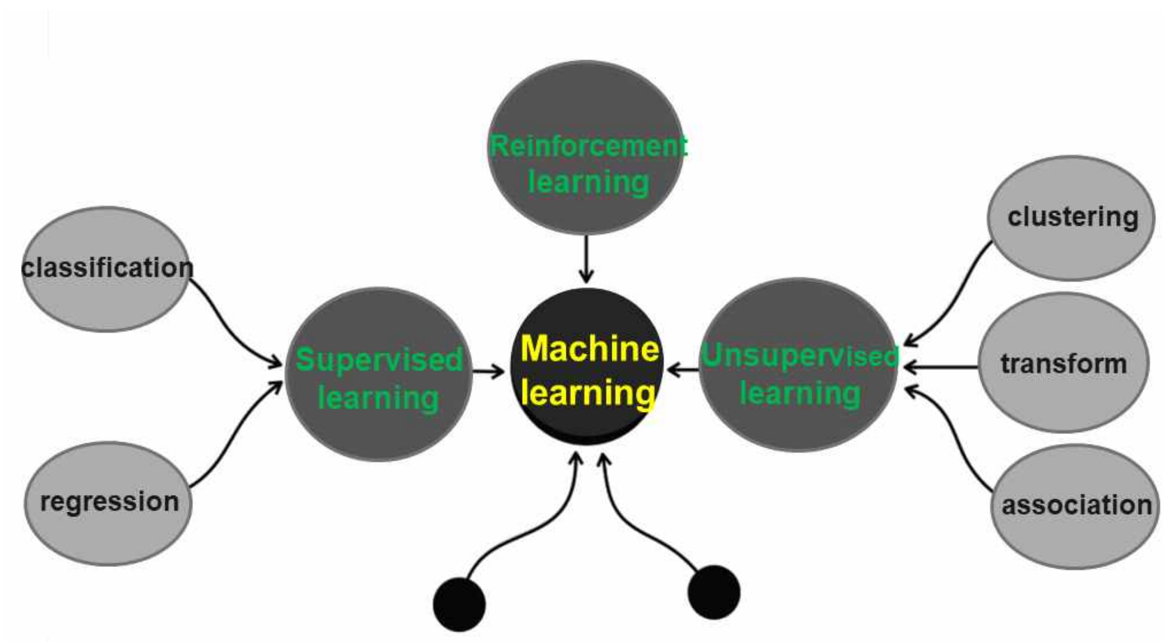


Figure 10. Classification of Machine Learning

Machine learning generally works in the following order:

1. Collect more than a certain amount of learning data.
2. Preprocess the collected data and find certain patterns and rules.
3. Learn and model with the patterns and rules found.
4. Evaluate the created model and make decisions and predictions.

As shown in Figure 10, machine learning algorithms are widely divided into three types based on the form of input information and data into the learning system.

- Supervised learning is to learn the mapping between input and output and apply when input and output pairs are given as data.
- Unsupervised learning is applied when there is only input and no output, and the goal is to find regularity between inputs. Unsupervised learning results are used as input to supervised learning or interpreted by human experts.
- Reinforcement learning is that the learner selects an action and affects the environment with the action and receives a reward through feedback and uses it as a guide for the learning algorithm. That is, it is applied to a system that takes an action corresponding to a given input, and examples of such a system include a robot or a player of a game. Unlike supervised learning, reinforcement learning does not give an output for a given input, that is, the correct answer behavior. Instead, rewards are given for the results of a series of actions, and the system uses these rewards to learn.

Supervised learning is a method of machine learning for inferring a function from training data. The training data usually includes the properties of the input object in the form of vectors and shows the desired result for each vector. Among these inferred functions, classification is indicating what kind of value a given input vector is, and the continuous value of output is called regression. The task of the supervised learning trainer is to correctly guess the value to be predicted for the given data from the training data. In order to achieve this goal, the supervised learning trainer should be able to generalize and handle situations that do not appear from the existing training data through an appropriate method. Figure 11 shows the overall process of the supervised learning process. First, when training data and test data for supervised learning training are prepared, features are extracted from the training data and a supervised learning model is selected to generate a learning model to be used. The accuracy of classification and regression results of the selected supervised learning model is evaluated by inputting test data into the generated model. Subsequently, instead of the test data, real data is inputted into the supervised learning model and used for classification and regression.

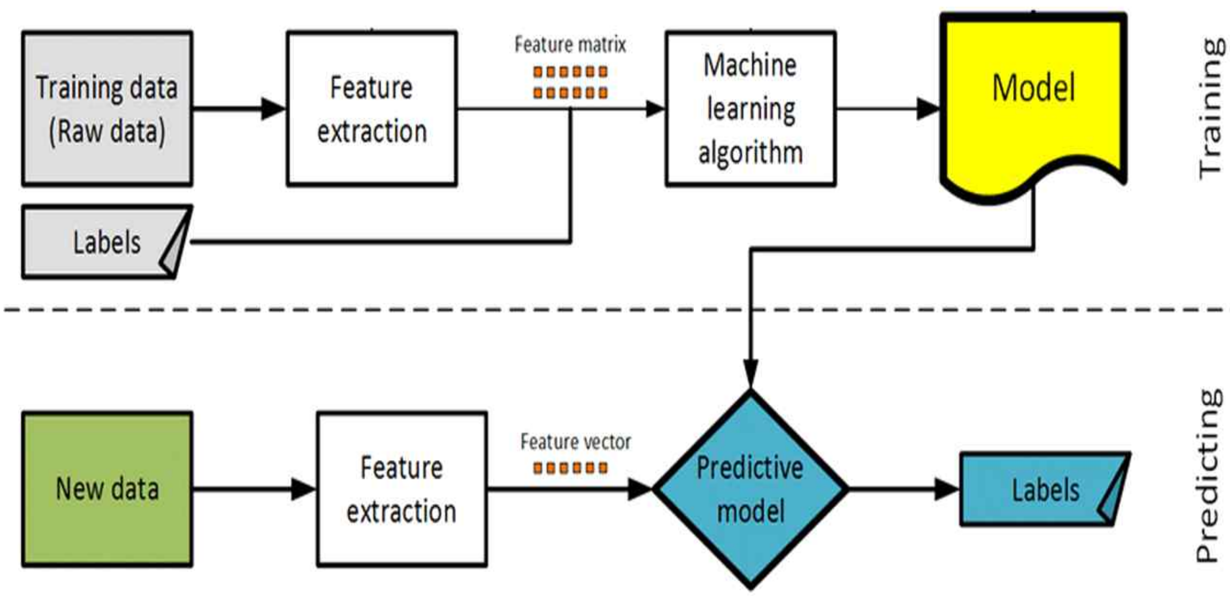


Figure 11. A process of Supervised machine learning model [38]

1) Classification

Classification is a method of finding out which group the newly input data belongs to after learning the training data in which the result value already exists. Classification is determined as one of the results of the data that has always been learned. If the result value is discrete, that is, the values of the result value may have 0,1,2,... Therefore, if it is finite, it is called a classification problem and it is one of the most common, well-researched, and generally most interesting issues in daily lives. Figure 12 shows 2 classes, circles and squares, and 2 lines, A and B. The model finds the best line A to set boundary between the classes. When new data is provided, it can divides the class where it belongs.

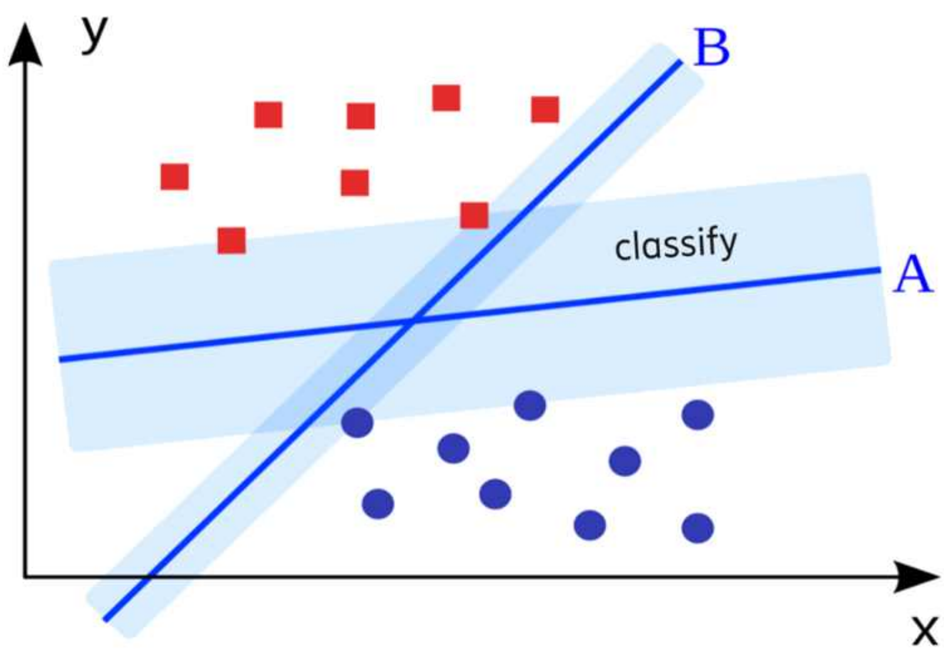


Figure 12. Example of Classification

2) Regression

Regression is called a regression problem when the result values are continuous, and when you want to draw a line or graph with the values that best explain the characteristics of certain data which is scattered, you can use the regression function to predict any pattern or trend. like classification, the answer does not have discrete result values like 0,1,2,.., but it can be predicted as a real value [39].

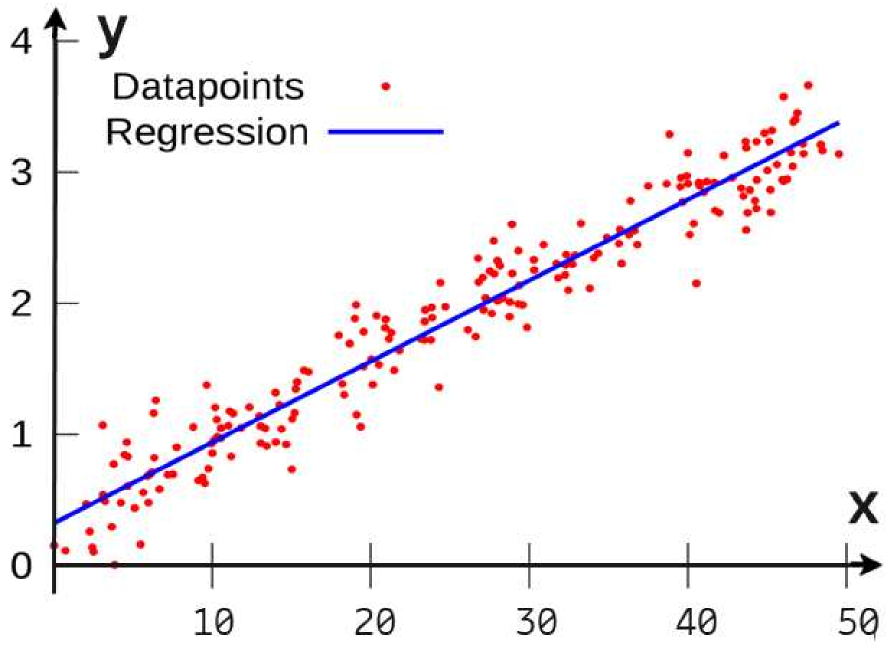


Figure 13. Example of Regression

The regression line, as shown in Figure 13, is the best fit line for a model, and main objective in this algorithm is to find this best fit line.

3) Supervised Learning Algorithm

Algorithms used for supervised learning, as shown in Figure 14, are k-Nearest Neighbors (KNN), Linear Regression, Discriminant Analysis, Support Vector Machines (SVM), Decision Trees, Naive Bayes, Ensemble Methods, Support Vector Regression (SVR), Gaussian Process Regression (GPR), Random Forest, and Neural Networks.

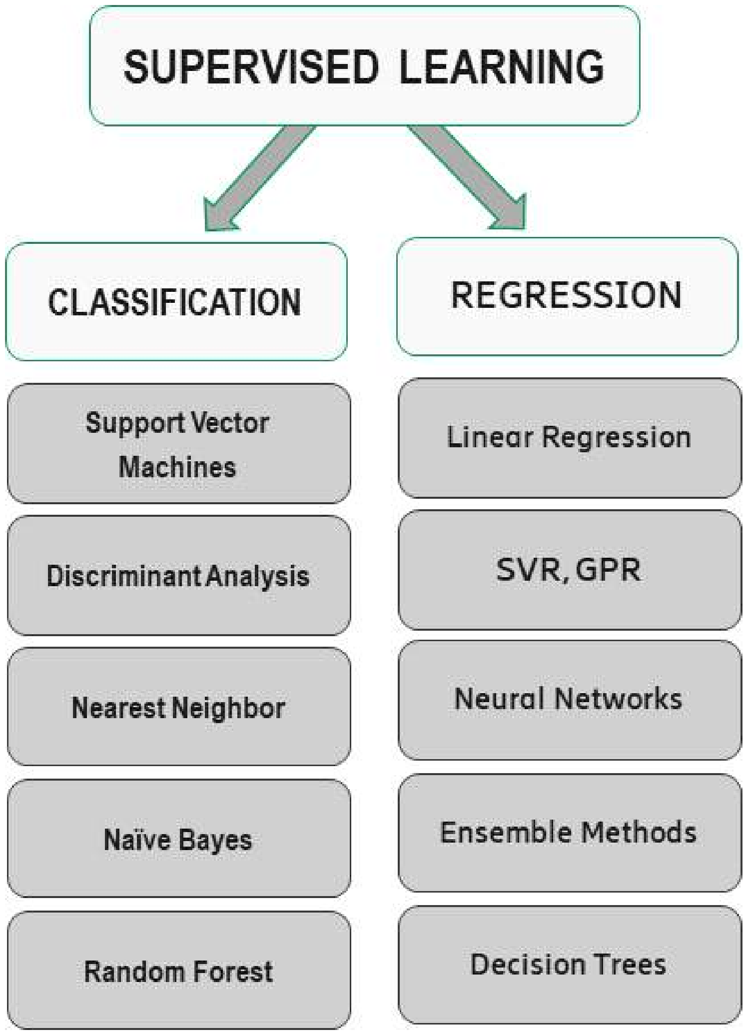


Figure 14. Types of Supervised Learning Algorithm

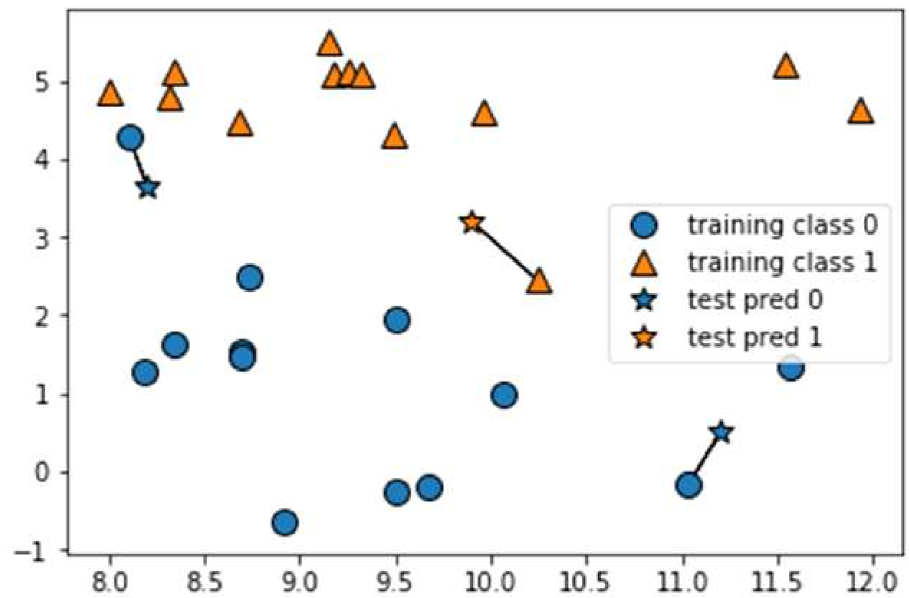
KNN(k-Nearest Neighbor)

k-Nearest Neighbor (KNN) is an algorithm proposed by Cover and Hart in 1968 [40, 41]. KNN is a method that determines the category of a new data point by selecting the K closest points from the training data and classifying the new data into one of the categories. A thing to consider in KNN is the measurement of the distance to the target point as a key part of the this algorithm, and the Euclidean distance measurement is normally used as a method of calculating distance. Treating all data columns, in the same way, can lead to errors owing to unexpected variables, therefore, it is necessary to discuss various distance calculation algorithms, such as subtracting the average distance for each category before summing the squares of the distances.

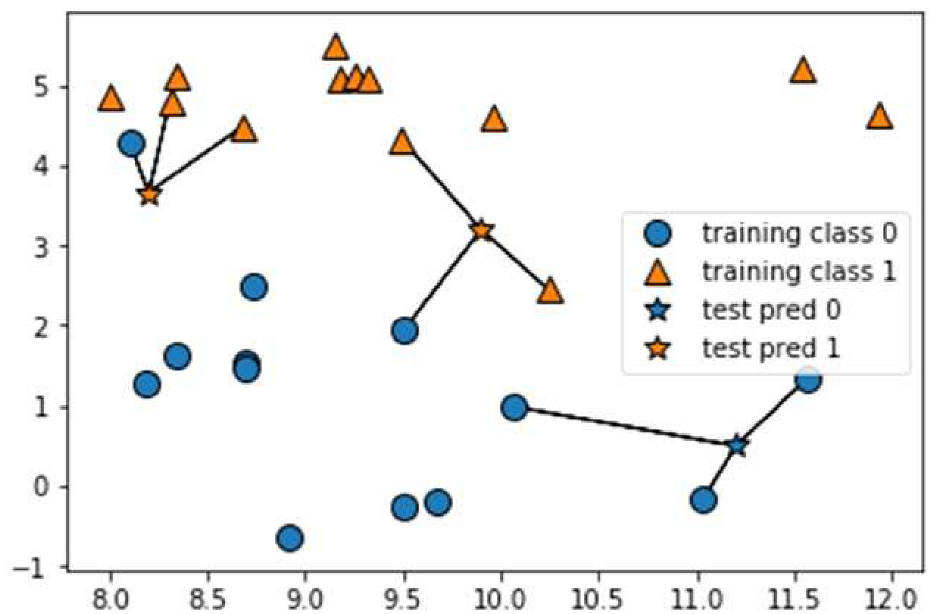
Table 3. Distance Metrics for KNN

Distance	Equation
Euclidean	$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$
Manhattan	$\sum_{i=1}^{k}\lvert x_i - y_i \rvert$
Minkowski	$\left(\sum_{i=1}^{k}(\lvert x_i - y_i \rvert)^q\right)^{1/q}$
Hamming	$\sum_{i=1}^{k}\lvert x_i - y_i \rvert$ $\quad x = y \Rightarrow D = 0$, $x \neq y \Rightarrow D = 1$

For example, as shown in Table 3, four distance metrics are considered to measure the closeness between two vectors x and y. For real data, the Euclidean, Manhattan, and Minkowski distance measurement method is used, and for data types, such as categorical or binary data, the Hamming distance measurement method is used [42].



(a) KNN classifier with k = 1



(b) KNN classifier with k = 3

Figure 15. Example of KNN

If the value of K is small, the the classification becomes sensitive to noise and overfitting occurs. Conversely, if the value of k is large, then underfitting occurs and computation becomes expensive.

Therefore, the K value should be selected appropriately based on the amount of data and the number of classes to be classified, as shown in Figure 15.

Decision Tree

A decision tree is used as a predictive model that connects the observed value and the target value for a certain item. It is one of the predictive modeling methods used in statistics, data mining, and machine learning. As one of the simplest classifiers, it has a very simple structure to graph a model using tools such as decision trees. This method is a model in which a final decision is made while selecting an appropriate node from the root, as shown in Figure 16 [43], in short, each node represents a feature, each branch represents a decision, and each leaf represents an outcome.

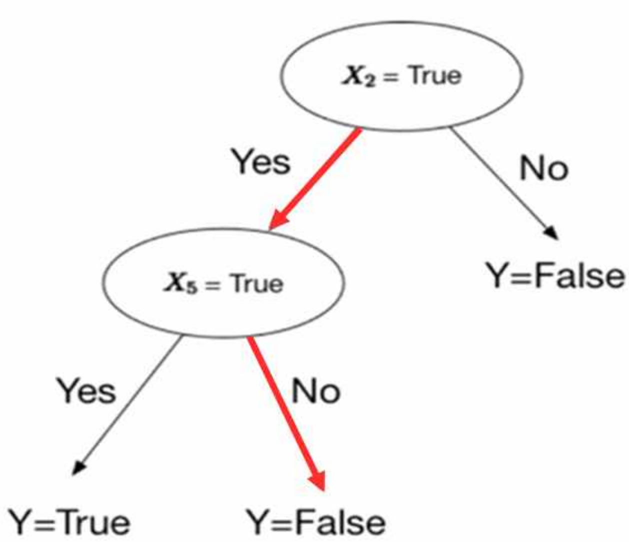


Figure 16. Example of Decision Tree

Random Forest

Random Forest utilizes ensemble learning method, which is a technique that combines many classifiers to provide solutions to complex problems and consists of many decision trees. This algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees, as shown in Figure 17. Increasing the number of trees increases the precision of the outcome [44].

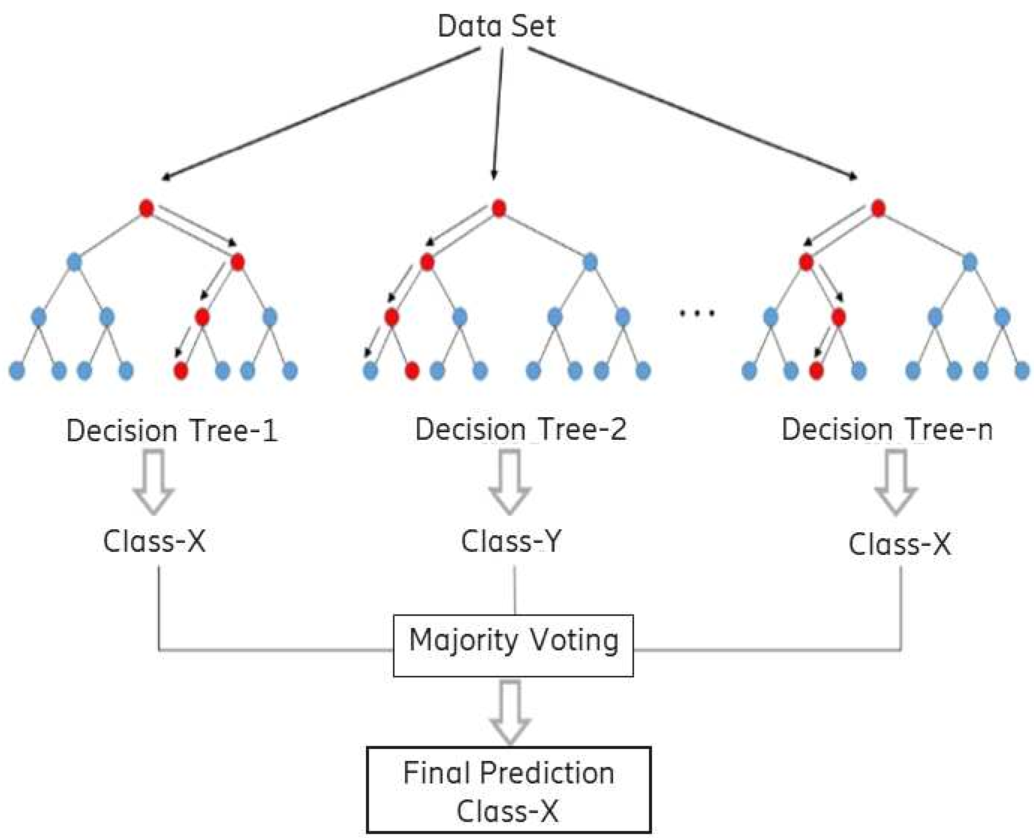


Figure 17. Example of Radom Forest

SVM(Support Vector Machine)

SVM (Support Vector Machine) is an algorithm that creates a hyper-plane with the maximum margin for each class to be classified from given data and classifies based on the hyper-plane when given new data. The margin means the distance from the data closest to the decision boundary among the training data to the decision boundary. The data closest to the decision boundary are the support vectors. When the training data are fixed, the margin and support vectors depend on the decision boundary. To reduce the generalization error, it is recommended to find the decision boundary with the maximum margin by maximizing the spacing between the two regions. In SVM, the linear decision boundary with maximum margin is called the hyper-plane, as shown in Figure 18. This finding of the maximum margin in the current dimension is called linear SVM [45].

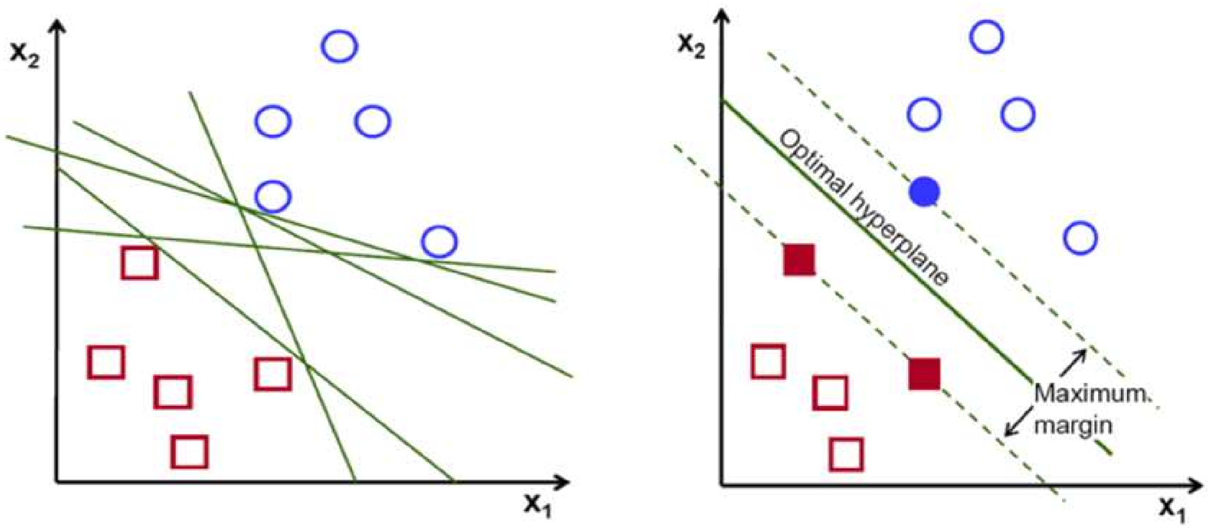


Figure 18. Example of SVM(Support Vector Machine)

D. Markov Process

The Markov process uses probability to model the manner in which a certain state changes over time. When the past states $(s_1, s_2, \ldots, s_{t-1})$ and the present state (s_t) are specified, it is assumed that the future state (s_{t+1}) is determined only by the present state; hence, the past state does not directly affect it. In other words, the present state is only affected by the past state at the immediate preceding stage and is not directly affected by the previous state.

In other words, the probability of the occurrence of the future state is the same both when the past and the present state are considered and when only the present state is considered [46, 47]. This can be expressed as follows:

$$P[s_{t+1}|s_t] = P[s_{t+1}|s_1, \ldots, s_t] \tag{1}$$

When a change occurs from state s at time t to the next state s' at time $t+1$, the state transition probability is expressed as follows:

$$Pss' = P[s_{t+1} = s'|s_t = s] \tag{2}$$

And the transition between states depends only on the previous state and it is a probabilistic model that determines the next state using the current state. Figure 19 is a graph with transition probabilities between states.

In addition, arranging the transition probabilities of all states in the form of a matrix is called a state transition probability matrix, as shown in Figure 20.

The Markov process with the state transition probability matrix converges to a stable probability distribution at a certain point in time through several transition processes. If the transition matrix does not change, then the probability of state does not change stably [48].

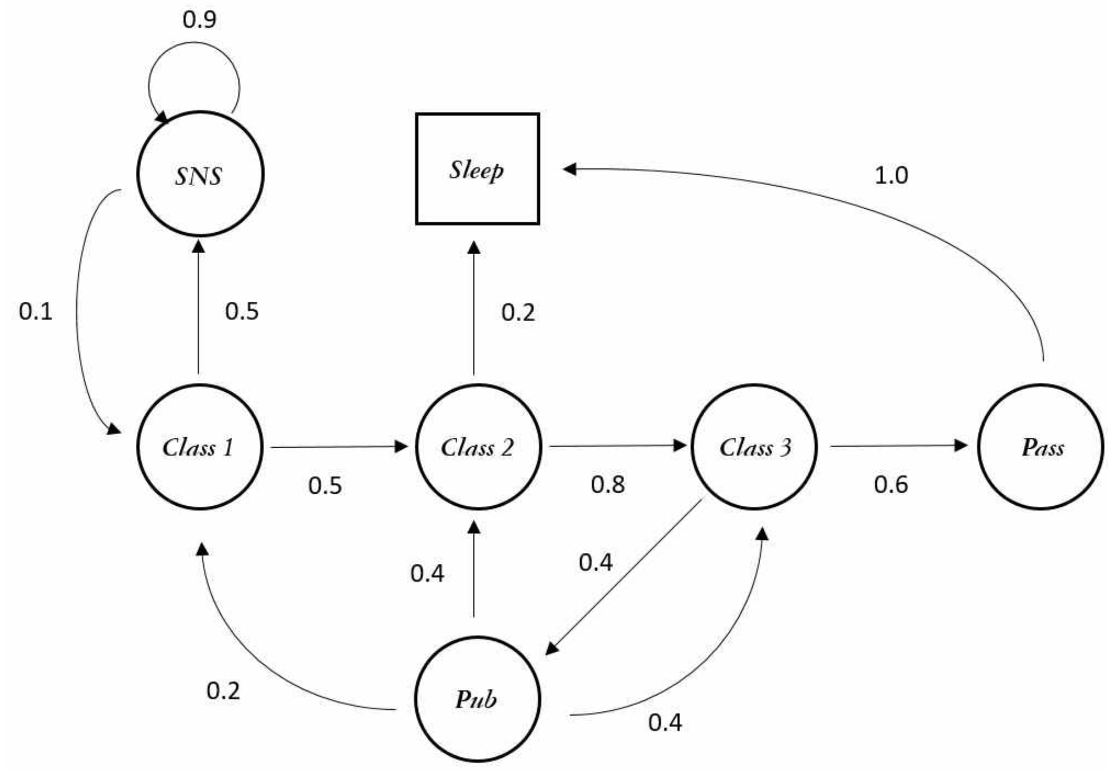


Figure 19. State Transition Diagram

$$
\mathcal{P} = \begin{array}{c} \\ \textit{Class 1} \\ \textit{Class 2} \\ \textit{Class 3} \\ \textit{SNS} \\ \textit{Pass} \\ \textit{Pub} \\ \textit{Sleep} \end{array}
\begin{array}{c}
\begin{array}{ccccccc} \textit{Class 1} & \textit{Class 2} & \textit{Class 3} & \textit{SNS} & \textit{Pass} & \textit{Pub} & \textit{Sleep} \end{array} \\
\begin{pmatrix}
 & 0.5 & & 0.5 & & & \\
 & & 0.8 & & & & 0.2 \\
 & & & & 0.6 & 0.4 & \\
0.1 & & & 0.9 & & & \\
 & & & & & & 1 \\
0.2 & 0.4 & 0.4 & & & & \\
 & & & & & & 1
\end{pmatrix}
\end{array}
$$

Figure 20. State Transition Probability Matrix

III. Proposed Machine Learning based Paging

A. Structure of Machine Learning based Paging

1. Algorithm of Machine Learning based Paging

In this thesis, a paging method using supervised machine learning and the Markov process is proposed for the 5G AMF. First, a UE classifier is implemented using KNN-supervised machine learning to classify subscribers using the UE based on the characteristics of their movement patterns. Meanwhile, UE movement data indicate that when UEs moves to adjacent gNodeBs, they detach and attach the previous gNodeB and new gNodeB, respectively, and these Registration are always collected in the AMF. To predict the current gNodeB of the UE, paging is performed based on the UE classification and the recent gNodeB transition statistics data extracted from the UE movement collected.

The dataset of the UE profile information used for supervised learning is as follows:

- First Registration Time: The time at which the first Registration in the AMF is completed.
- Last Registration Time: The time at which the last Registration/Service is performed in the AMF.
- Last Update Type: The last Registration Type of UE in the AMF.
- Last Service Type: The last Service Type of UE in the AMF.
- gNodeB ID/TAC: The UE's latest gNodeB ID/TAC information.
- Latest TA List: The UE's latest TA list information.

Table 4. UE Profile Dataset

No.	Item	Example
1	First Registration Time	MM-DD hh:mm:ss
2	Last Registration Time	MM-DD hh:mm:ss
3	Last Update Type	Periodic/Mobility
4	Last Service Type	UE-Initiated /Network-Initiated
5	gNodeB ID	1556
6	TAC	10154
7	Latest TA List	240-81-10154, 240-81-10133, 240-81-21001 …

This thesis consider classifiers such as the KNN, Random Forest, Decision Tree, and SVM applied with the UE profile information registered to the AMF, as shown in Table 4.

In this thesis, Because 5G services has not been commercialized, the UE profile information for the 5G AMF could not be obtained. Therefore, the UE profile information obtained in 4G MME was converted to satisfy the 5G technical specifications and used in experiment.

Subscribers were classified into two UE groups, as shown in Table 5. One UE group included local residents and office workers who did not move significantly daily, and the other UE group included travelers, non-office workers, etc. This thesis expect the overall paging performance to be improved by classifying the UE with the supervised-learning-based UE classification and applying the weight to the confidence level used for probabilistic gNodeB selection during the paging process.

For the experiment involving four classification models, I.e, KNN, Random Forest, Decision Tree, and SVM, 1000 UE profile information of every 10 minutes cycle for four days was collected for supervised learning, and 1000 UE profile information for one day was used for the performance evaluation of the four models [49].

Performance evaluation of the classification models was conducted using scikit-learn, a machine learning library for the Python programming language. Subsequently, the test results of classification were compared [50].

Figure 21 shows the test results of the classification accuracy. It was observed that both the KNN and SVM, which are candidates for UE classification model, provided satisfactory classification accuracy. In addition, the simplicity and convenience of implementing the supervised learning classifier model in the AMF were considered during model selection. In my study, the KNN model for classification was selected owing to its high accuracy and simple implementation. The K value and distance metric in the KNN were used for 10 and Euclidean distance measurement, respectively. The recent TA list that was weighted based on the distance from the last gNodeB and TA, and the time stamp converted into the duration information in the UE profile information were used to measure the Euclidean distance.

Table 5. UE Group Category

UE Group	Type	Remark
1	Stable UE	Resident, Office worker
2	Random UE	Traveler, Non-office worker

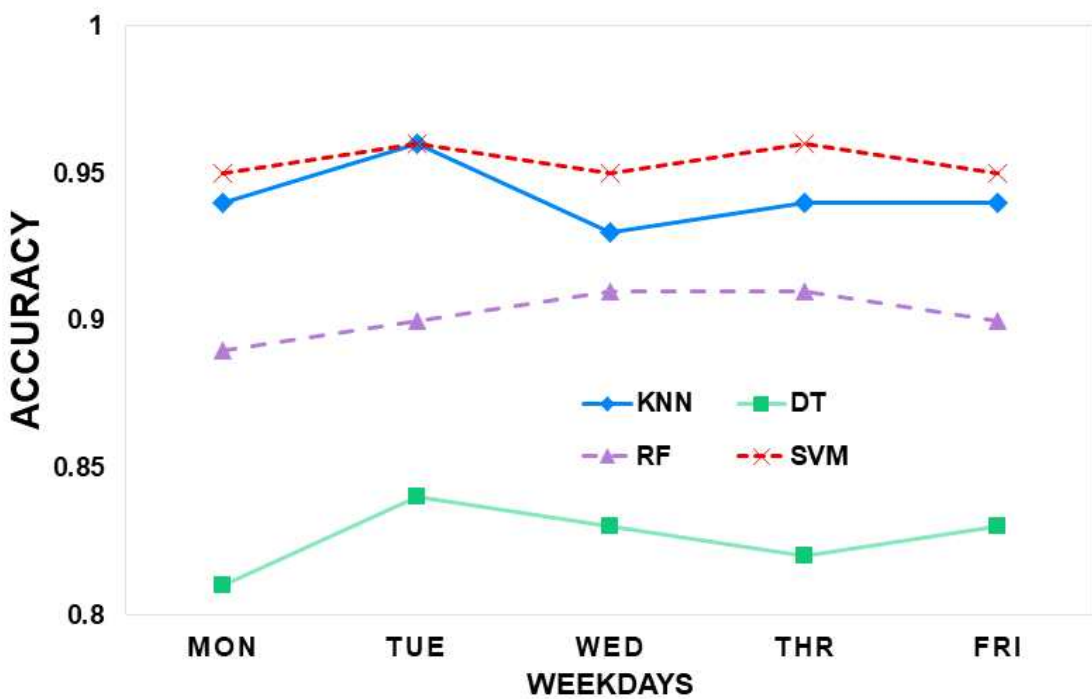


Figure 21. Classification Accuracy

As shown in Figure 22, the proposed probabilistic paging algorithm applied with machine learning in the AMF is categorized into two functions, as follows:

The first function collects the UE profile information in the AMF and conducts a UE classifier model applied with machine learning. In the paging experiment, when paging is attempted toward the target UE, the weight ratio of the UE Group, which is obtained by performing the UE classifier model, is used to select appropriate gNodeB.

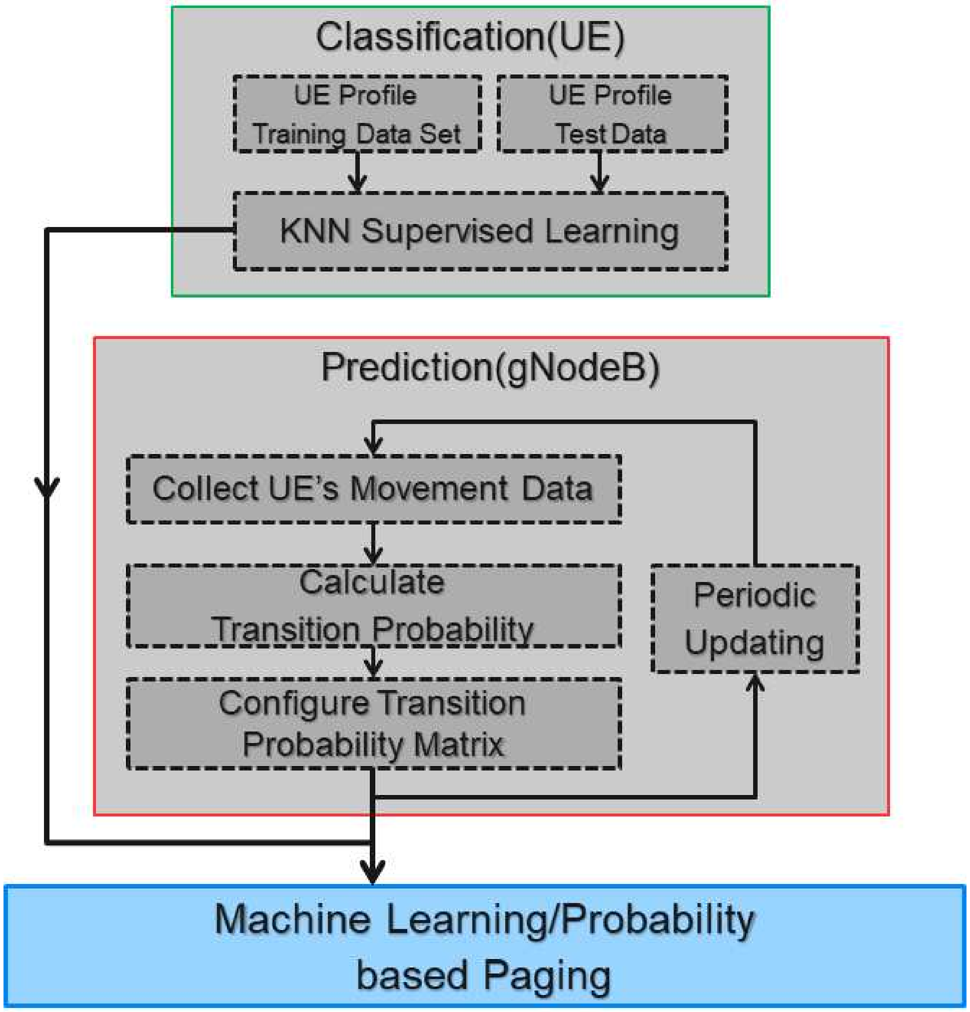


Figure 22. Algorithm of the proposed paging

The second function collects the UE movement around the gNodeBs and then classifies the acquired UE movement by the gNodeB. The transition probability for each gNodeB is sorted in the descending order to perform probabilistic paging method.

By periodically performing these processes at the AMF, each gNodeB updates the transition probability using the latest UE movement data that have been changed recently.

The paging method with KNN-based UE classification and the UE movement prediction to gNodeB is performed by substituting the second paging range of the paging range profile, as shown in Figure 23.

Therefore, when the AMF uses the gNodeB list for the second step of paging, the paging range is selected by the UE classifier applied with the KNN and the probabilistic gNodeB list proposed herein, then paging was attempted.

When the proposed paging is not activated, as shown in Figure 23, the AMF attempts paging based on the latest visited gNodeB list in terms of the second paging range.

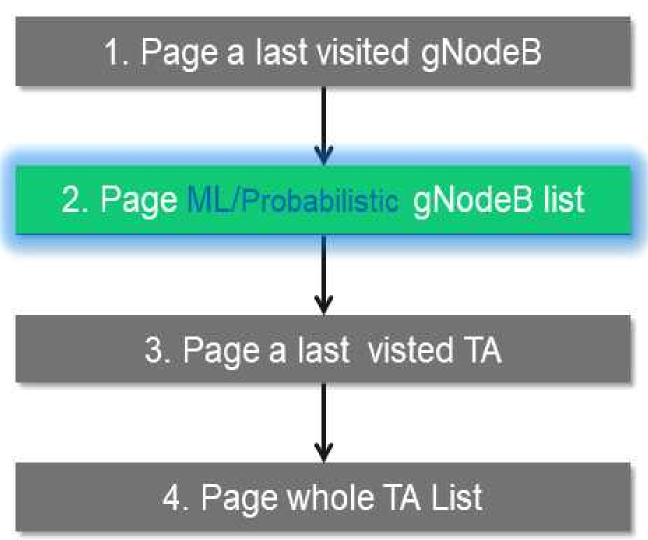


Figure 23. proposed AMF Paging Range profile

2. Operation of Machine Learning based Paging

Figure 24 shows the internal block structure and operation procedure on the AP (Application Process) and CP (Central Process) for the probabilistic paging method applied with machine learning in the AMF.

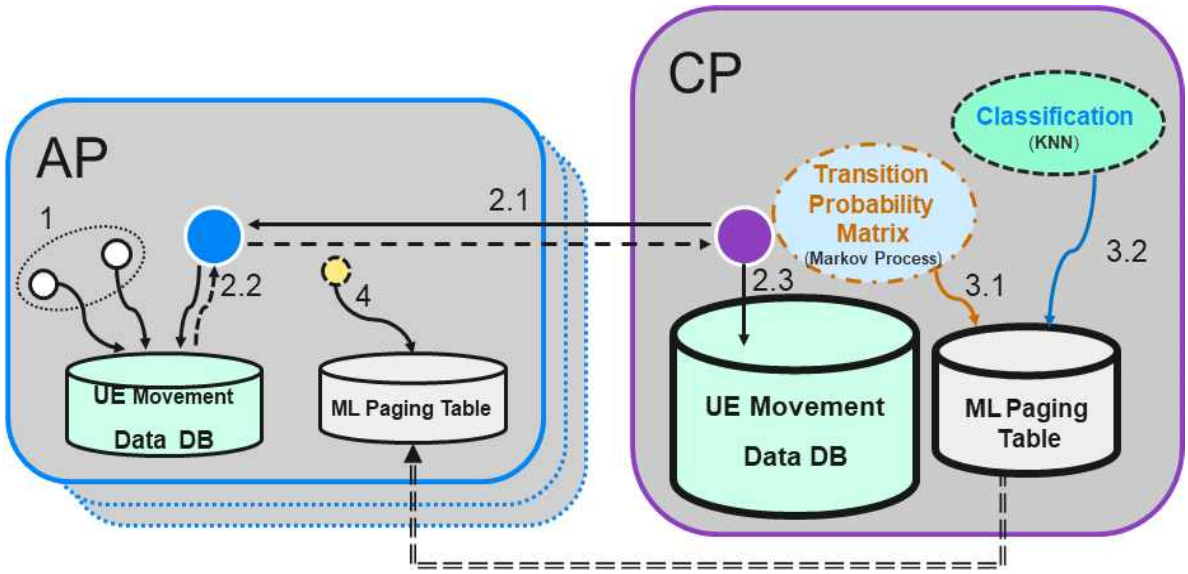


Figure 24. Structure and process of AMF for proposed paging

CP : Periodically collects, accumulates, and stores the UE movement data of the gNodeB that is sent from each AP. After calculating transition probability information using the newly updated UE movement data, subsequently, the result is delivered to each AP. And a UE classifier model based on the KNN-supervised learning using the UE profile information is also delivered to each AP.

AP : Collects the UE movement data from the gNodeB and the UE profile information, and performs paging using the Machine Learning Paging Table that has the probabilistic gNodeB list and the weight ratio of the UE classifier model applied with machine learning.

UE Movement Data DB : Stores the UE movement data and UE profile information collected from all gNodeBs, and it exists in both AP and CP.

Machine Learning Paging Table : Stores the transition probability information of each gNodeB that is newly calculated using the UE movement data from the gNodeB, and is updated periodically and exists in CP and AP. Also, it stores a UE classifier model based on KNN-supervised learning using the UE profile information, and is used for the probabilistic paging method applied with machine learning.

As shown in Figure 24, which describes the structure and process between internal blocks within the AMF,

In step 1, when the UE location information changes owing to the UE movement from the gNodeBs, or Service provision in each AP of the AMF. The AMF collects the UE movement information from the gNodeB and stores in the UE Movement Data DB, and then the UE profile information is updated and stored.

In step 2, the AMF collects all data from the UE Movement Data DB in which the CP stores each gNodeB movement information for a duration from all APs. Subsequently, when all the collected gNodeBs are classified and delivered to the central processing unit of the CP, the central processing unit adds the newly collected gNodeB information to the cumulative gNodeB which is stored in the previous duration and updates the accumulated information for all gNodeBs and stores in the UE Movement Data DB.

In step 3, new transition probability information by Markov process are calculated for each gNodeB based on the updated cumulative information of the gNodeB, and the updated probabilistic gNodeB list is stored and delivered to the Machine Learning Paging Table in each AP. In addition, the UE classifier model that has performed the KNN supervised-learning-based on the UE profile information is also delivered and stored into the Machine Learning Paging Table.

Step 4, when the probabilistic paging method applied with machine learning is activated in the paging profile configuration of the AMF, the probabilistic paging method applied with machine learning uses the gNodeB transition probability information and the UE classifier model, to select the optimal gNodeB in the Machine Learning Paging Table.

B. Process of Machine Learning based Paging

The AMF selects the final gNodeBs from the probabilistic gNodeB list for paging based on a combination of three variables: the weight ratio of the UE group in the UE classifier model result, the transition probability of the gNodeB, and the confidence level in the paging profile configuration. During the paging, the weight corresponding to each the UE group is applied to the UE when probabilistic gNodeBs are selected. In this thesis, the AMF classifed the UE into two UE groups, as show in Table 5, and collects the UE profile information of these UE groups.

To predict the UE movement to the gNodeB, the AMF periodically collects a significant amount of UE movement data between gNodeBs and the UE profile information from the Registration procedure based on the UE mobility events. For example, the Periodic or Mobility Registration Update, Service Request, and Handover Procedure contain the latest UE movement data. After collecting movement data between gNodeBs, the AMF periodically combines with the newly collected UE movement data. The training data were obtained by multiplying the decay factor for past UE movement data at the current point and calculates the transition probability of moving from one gNodeB to all of its adjacent gNodeBs.

The confidence level of the paging profile configuration is a set of configurable probability values and is a threshold that terminates the process of selecting the gNodeB for the paging process. A confidence level includes one or more probability values for configuring a confidence level with more than one probability value that provides the backup confidence level values for retries, even if the first paging attempt fails. Therefore, the number of probability values of the confidence level should be equal to the number of paging attempts configured for each paging step in the AMF paging profile configuration. This is because the number of probability values of the confidence level is the same as the number of paging attempts. The confidence level may be set in units of the entire AMF or the paging profile configuration. When both are set, the confidence level of the paging profile configuration is prioritized.

Figure 25 shows an example of the probabilistic gNodeB paging applied with machine learning, where the AMF selects gNodeB2 with the highest transition probability based on the last visited gNodeB0, then gNodeB1 with the second highest, followed by gNodeB4 with the next highest. In other words, the gNodeBs are selected to perform paging in the descending order of transition probability values. When the sum of the accumulated transition probabilities of the selected gNodeBs satisfies or exceeds the probability value of the specified confidence level, no further gNodeBs are selected for the paging attempts. The paging range profile shows that the AMF always pages to the last visited gNodeB0, along with the selected gNodeBs, in every paging attempt.

The example shown in Figure 25 illustrates the manner by which the AMF performs paging based on the probabilistic gNodeB list applied using machine learning.

The AMF analyzed the movement data of UEs in the last visited gNodeB0 and calculated the transition probabilities of four adjacent gNodeBs(gNodeB1, gNodeB2, gNodeB3, and gNodeB4) from gNodeB0, as shown in Figure 25, and the transition probabilities were sorted in the descending order [51].

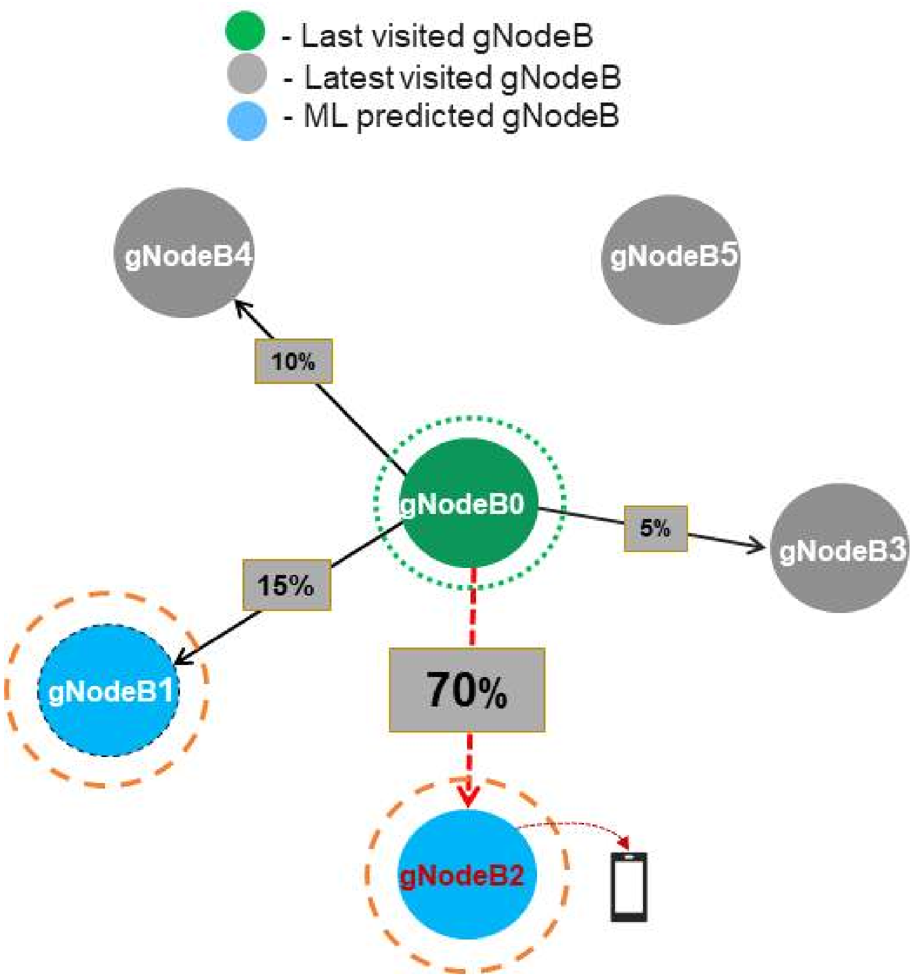


Figure 25. Example of the proposed paging

The AMF selects the gNodeB to perform paging in the order from the gNodeB with the highest to lowest transition probability. Table 6 lists the state transition probabilities for each gNodeB based on the UE movement from gNodeB0. The total number of the UE that moved from gNodeB0 to the neighboring gNodeB was 2000. The number of UE that moved to gNodeB2 was 1400 with 70% transition probability, gNodeB1 was 300 with 15% transition probability, gNodeB4 was 200 with 10% transition probability, and gNodeB3 was 100 with 5% transition probability.

Table 6. Example of State Transition Probabilities for gNodeB

UE movm'nt Path (from gNodeB0)	UE movm'nt Count	Transition Probability
gNodeB2	1400	0.7(70%)
gNodeB1	300	0.15(15%)
gNodeB4	200	0.1(10%)
gNodeB3	100	0.05(5%)

When No. 4 of the paging profile configuration in Table 7 was selected, the number of paging attempts was set to three times, and the confidence level for the node criterion was set to three probability values: 0.7, 0.8, and 0.9. Subsequently, the AMF performs three paging attempts in the following second machine-learning-based probabilistic paging.

In step 1 of paging, the AMF first selects gNodeB2. In this regard, the transition probability of gNodeB2 is 0.7, which is equal to or greater than the first probability value (0.7) at the confidence level. Next, when the target UE is classified into UE group 2 by the UE classifier model, the gNodeB is assigned the corresponding weight ratio (100%), as show in Table 8. Therefore, the AMF attempts to page gNodeB2 and gNodeB0. Also, the AMF always performs paging including the last visited gNodeB0 in every step.

In step 2 of paging, if step 1 of paging fails, the AMF selects gNodeB2 and gNodeB1, where the transition probabilities of the two gNodeBs are 0.85, which is equal to or greater than the second probability value (0.8) in the confidence level. Next, when the target UE is classified into UE group 1 by the UE classifier model, the gNodeB is assigned a corresponding weight ratio (50%), as shown in Table 8. Therefore, the AMF finally pages gNodeB0 along with gNodeB2; however, by applying a weight ratio of 50% to the gNodeBs, gNodeB1 is excluded from the paging attempts.

Table 7. Example of Changing AMF Paging Profile Configuration

Paging Profile	Last Visited gNodeB	Latest Visited gNodeB List / ML/Probabilistic gNodeB List	Last Visited TA	TA List	Confidence Level
1	0	0	0	4	-
2	0	0	2	3	-
3	0	1	2	2	-
4	0	3	2	1	0.7, 0.8, 0.9
...	...	...	...	...	...
20	3	2	2	2	-

In step 3 of paging, if the step 2 of paging fails, then the AMF selects gNodeB2, gNodeB1, and gNodeB4, where the sum of the transition probabilities of the four gNodeBs is 0.95, which is equal to or greater than the third probability value (0.9) at the confidence level. Next, because the target UE has already been classified as UE group 1 by the UE classifier model, the AMF selects the gNodeBs with the weight ratio of 50%. However, the AMF attempts to page gNodeB2, gNodeB1, and gNodeB4 together with gNodeB0. Because the AMF has a history of the paging failure to which a weight ratio is previously applied, the weight ratio (100%) is applied instead of the weight ratio (50%) of UE group 1 in the UE classifier model.

Table 8. Example of Weight Table of UE Group

UE Group	Weight(%)	Remark
1	50	Resident, Office worker
2	100	Traveler, Non-office worker

In another example, i.e., No. 20 of the paging profile configuration in Table 7, when applying the second machine learning probabilistic paging, the three probability values of the confidence level for the node criteria in the paging profile configuration were 0.7, 0.8, and 0.9; however, if the number of paging attempts (2) is less than the number of probability values of the confidence level 3, then the AMF selects the lowest probability value from the paging profile configuration and performs paging attempts. When the second paging attempt is completed, the confidence level of the remaining high probability value (0.9) is disregarded. Conversely, if the number of paging attempts in the paging profile configuration is greater than the number of probability values of the confidence level, then the AMF performs paging attempts from the lowest probability value. When paging attempts using all the probability values are completed, the AMF reuses the highest probability value of the confidence level and performs the remaining paging attempts.

The proposed paging system excludes the gNodeB with the lowest transition probability among the candidates selected from the gNodeB list based on the transition probability and confidence level of the gNodeB.

After machine learning-based probabilistic paging is activated, the number of paging attempts, confidence level of the paging profile configuration, weight assignment of the UE classifier model, and period of updating and calculating the probabilistic value based on the collected UE movement data should be optimized repeatedly to obtain the best gNodeB paging performance results.

IV. Experimental Results and Discussion

A. Experimental Environment and Scenario

For the performance evaluation of the probabilistic paging method applied with machine learning, the AMF system in which the KNN supervised-learning-based UE classifier model, Markov-process-based probabilistic paging method, UE, gNodeB, and SMF are implemented is shown in Figure 26.

The experimental environment was implemented by building a VM-based 5G simulator with scripts such as Periodic or Mobility Registration Update procedure, Service Request procedure, and paging process.

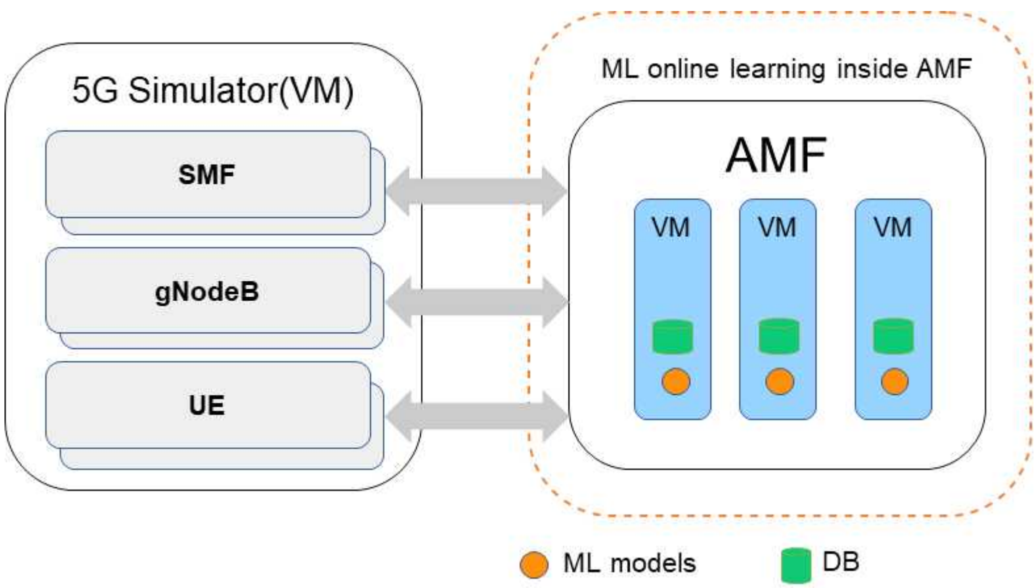


Figure 26. Experiment Environment for the proposed paging

In addition, UE movement data of approximately 200000 UEs and approximately 1300 eNodeBs were collected from a commercial MME system for one week. Subsequently, they were fitted and adjusted for the experimental environment of the standard 5G specification. in addition, the KNN-supervised learning-based UE classifier model was conducted using the UE profile information of the 1000 subscribers for 1 week.

The UE group of subscribers was classified into UE group 1, which included local residents and office workers who moved less between gNodeBs, and the remaining was classified as UE group 2. The setting for all the gNodeBs and UE were organized in the order of time in the 5G simulator, whereas the gNodeB list and transition probability based on the proposed method were updated every 20 minutes.

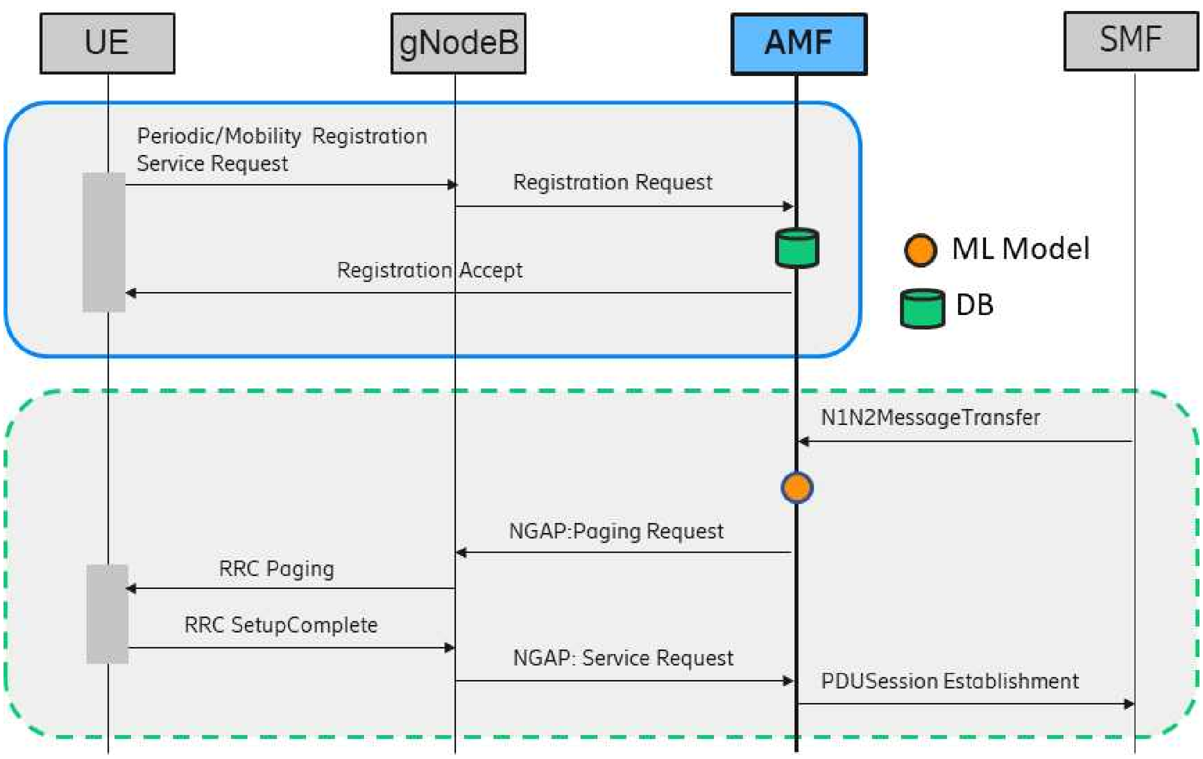


Figure 27. Experiment Scenario for the proposed paging system

Figure 27 shows the experimental scenario. When the UE sends a periodic or mobility registration and service request message to the AMF, the AMF stores the UE movement data in the gNodeB DB, generates the UE classifier model, and updates the machine-learning paging table. When the SMF sends a service request message to the AMF, the AMF verifies the UE state and performs probabilistic paging with machine learning.

B. Experimental Results and Evaluation

The paging profile configuration of the AMF used in the experiment environment is as follows: The typical paging approach was implemented via only one attempt from the first paging range with the last visited gNodeB to the fourth paging range with TA list paging. Meanwhile, the proposed machine-learning-based probabilistic paging did not attempt the first paging with last visited gNodeB, and the second paging based on the probabilistic paging method applied with machine learning was attempted with three times. The confidence level of the paging profile configuration was set to 0.7, 0.8, and 0.9, and the weight ratio of UE groups 1 and 2 of the KNN-supervised-learning-based UE classifier model was set to 50% and 100%, respectively, after which the statistics regarding the paging signals were obtained from the AMF. Subsequently, the experimental results were compared with each other. For the performance evaluation, three cases were performed: typical paging, probabilistic paging applied with Markov process, and probabilistic paging with the KNN supervised-learning-based UE classifier model and Markov process.

Figure 28 shows the change in the overall paging signal. As shown in Figure 28, when probabilistic paging was conducted, the total number of paging signals in the AMF decreased by 60% as compared with the typical paging shown in Figure 8, because the remaining paging steps of the paging profile configuration were not performed after the success of the paging. When the proposed probabilistic paging with machine learning was applied, the total number of AMF paging signals reduced by 70% compared with that of the typical paging.

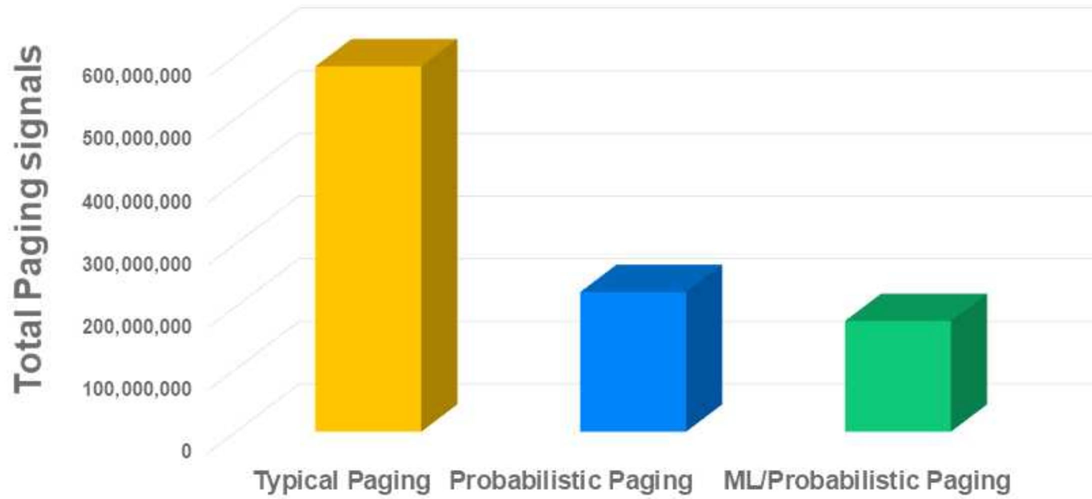


Figure 28. Total paging signals observed under the different conditions of paging system

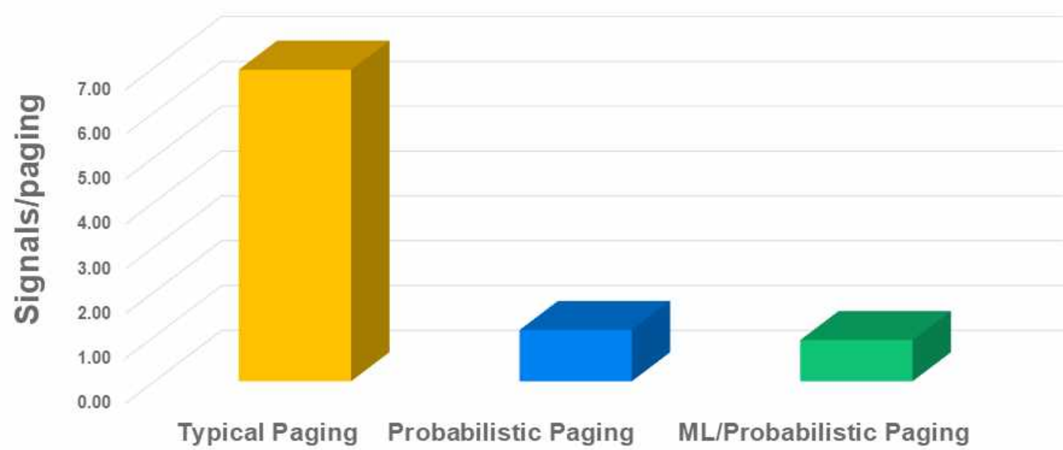


Figure 29. Average paging signals observed under the different conditions of paging system

In terms of the average number of signals per paging shown in Figure 29, the probabilistic paging method decreased it by 83%, whereas the probabilistic paging method with machine learning decreased it by 86%.

Based on results of two previous experiments, It was observed that the UE classifier model classified the UE into groups based on the subscriber type and assigned a weight ratio to select the best gNodeB. This reduces the overall paging signal and contributes positively to the AMF performance. In this experiment, all paging attempts were performed up to the entire TA list of the last step; therefore, the final paging success rate was the same for both the existing gNodeB list paging method and the proposed machine-learning-based probabilistic paging method.

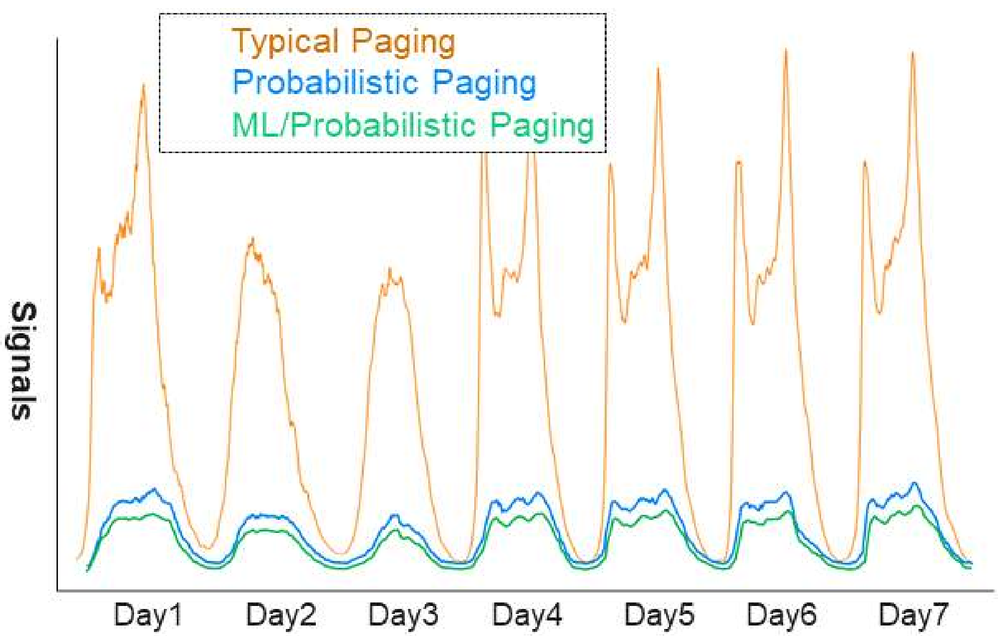


Figure 30. Paging signals per second observed under the different conditions of paging system

Figure 30 shows the number of paging signals per second over a week, As shown, the number of paging signals per second reduced by an average of 40%-80% per day. In addition, the probabilistic paging method applied with machine learning showed that the number of paging signals per second reduced by up to 25% owing to the probabilistic paging method involving a Markov process.

V. Conclusion

In this thesis, the structure and operation procedure of the probabilistic paging method applied with machine learning were proposed to reduce the number of paging signals in the 5G AMF system, and the performance was evaluated based on experiments on a 5G simulator and the AMF.

The proposed paging method reduced the average number of signals per paging and the total number of paging signals in the AMF compared with the typical paging method.

The experimental results confirmed that it contributed to the improvement in the operational stability and performance of the AMF system.

Hence, the proposed probabilistic paging method applied with KNN machine learning is a supervised learning model with a relatively simple structure; furthermore, it is applicable to the commercial operating AMF system for predicting resource allocation and managing abnormal problems in 4G/5G core networks in advance.

References

[1] Ericsson. Ericsson Mobility Report. [Online]. Available: https://www.ericsson.com/49da93/assets/local/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf, Jun. 2020.

[2] A. Gupta and R. K. Jha, “A Survey of 5G Network: Architecture and Emerging Technologies,” IEEE Access, vol. 3, pp. 1206-1232, Jul. 2015.

[3] 3rd Generation Partnership Project(3GPP) TS 38.413, “Next Generation Radio Access Network (NG-RAN); NG Application Protocol (NG-AP),” Jan. 2020.

[4] L. O. Widaa and S. M. Sharif, “Effect of Tracking Area List overlapping in reducing overall signaling overhead in Long Term Evolution system,” 2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE), pp. 397-400, Sep. 2015.

[5] Larsson, F. and Karlsson, A., “Evaluation of Machine Learning Algorithms to Reduce Paging Signalling in a Telecom Network,” pp. 89, Dec. 2017.

[6] 3rd Generation Partnership Project(3GPP) TS 23.502, “Procedures for the 5G system(5GS),” Jun. 2020.

[7] S. Sivasankar and R. Challa, “Closed Loop Paging Optimization for Efficient Mobility Management,” IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), pp. 1-7, Jan. 2021.

[8] ITU, “Setting the Scene for 5G: Opportunities & Challenges” 2018.

[9] Kim, Bumsu, “ICT-Based Business Communication with Customers in the 4th Industrial Revolution Era,” Business Communication Research and Practice. vol. 2, pp. 55-61, Oct. 2019.

[10] NGMN Alliance, “NGMN 5G White Paper,” Next Generation Mobile Networks, White paper, Feb. 2015.

[11] 5G-PPP View on 5G Architecture (Version 2.0) white paper:

https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017_For-Public-Consultation.pdf, Jul. 2017.

[12] P. Popovski, K. F. Trillingsgaard, O. Simeone and G. Durisi, “5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View,” IEEE Access, vol. 6, pp. 55765-55779, Sep. 2018.

[13] 3rd Generation Partnership Project(3GPP), “5G-NR Workplan for eMBB,” 2017.

[14] M. Kim, J. Park, M. Na, S. Jo, “Development Direction of 5G Mobile Communication Technology,” KICS Information & Communication Magazine, vol. 32, no. 9, pp. 46-54, Sep. 2015.

[15] Verizon Wireless, “5G Ultra Wideband Wireless Home Network,” https://www.verizonwireless.com/5g/home/, 2019.

[16] C. Bockelmann, “Towards Massive Connectivity Support for Scalable mMTC Communications in 5G Networks,” IEEE Access, vol. 6, pp. 28969-28992, May 2018.

[17] Qi Zhang and Frank H.P. Fitzek. “Mission Critical IoT Communication in 5G,” Future Access Enablers for Ubiquitous and Intelligent Infrastructures. FABULOUS 2015. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, vol. 159, pp. 35-41, Sep. 2015.

[18] 3rd Generation Partnership Project(3GPP) RP-191561, “Support of NR Industrial Internet of Things(IoT),” 2019.

[19] 5G Americas White paper, “New Services and Applications with 5G Ultra-Reliable Low Latency Communications,” https://www.5gamericas.org/wp-content/uploads/2019/07/5G_Americas_URLLLC_White_Paper_Final__updateJW.pdf, Nov. 2018.

[20] 3GPP RP-191584, “Physical layer enhancements for NR URLLC,” 2019.

[21] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. M. Leung, ‘‘Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges,’’ IEEE Commun. Mag., vol. 55, no. 8, pp. 138–145, Aug. 2017.

[22] P. Ashwood-Smith, “Why end-to-end network slicing will be important for 5G”, ITUNews, Jul. 2017.

[23] Por Town Kassis, “NSA or SA? Who is the real 5G network mode for smartphones and what's the difference?,” www.gearbest.com, Oct. 2019.

[24] Jihoon Lee, “Core Network Migration Paths to 5G,” www.netmanias.com, May 2019.

[25] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang and N. Li, “5G Deployment: Standalone vs. Non-Standalone from the Operator Perspective,” IEEE Communications Magazine, vol. 58, no. 11, pp. 83-89, Nov. 2020.

[26] Karim Rabie, “Core Network Evolution - 5G Service based Architecture,” www.netmanias.com, Dec. 2017.

[27] 3rd Generation Partnership Project(3GPP) TS 24.501, “Non-Access-Stratum (NAS) protocol for 5G System(5GS); Stage 3,” Dec. 2020.

[28] 3rd Generation Partnership Project(3GPP) TS 23.501, “System architecture for the 5G System(5GS),” Jun. 2021.

[29] R.-H. Liou, Y.-B. Lin, and S.-C. Tsai, “An Investigation on LTE Mobility Management,” IEEE Transactions on Mobile Computing, vol. 12, no. 1, pp. 166–176, Jan. 2013.

[30] 3rd Generation Partnership Project(3GPP) TS 36.300, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description,” Apr. 2017.

[31] 3rd Generation Partnership Project(3GPP) TS 38.300, “NR and NG-RAN Overall description,” Jan. 2020.

[32] 3rd Generation Partnership Project(3GPP) TS 29.500, “5G System; Technical Realization of Service Based Architecture,” Jul. 2020.

[33] 3rd Generation Partnership Project(3GPP) TS 38.331, “NR; Radio Resource Control (RRC); Protocol specification,” Dec. 2019.

[34] Y. Xiao, H. Chen, X. Du and M. Guizani, “Performance Analysis of Blanket Paging, Sequential Probability Paging, and Pipeline Probability Paging for Wireless Systems,” IEEE Transactions on Vehicular Technology, vol. 56, no. 5, pp. 2745-2755, Sep. 2007.

[35] Arthur L Samuel, “Some Studies in Machine Learning Using the Game of Checkers,” IBM Journal of Research and Development 3, pp. 210-229, Jul. 1959.

[36] T. Mitchell, “Machine Learning,” McGraw Hill, 1997.

[37] S. Shalev-Shwartz, S. Ben-David, “Understanding Machine Learning: From Theory to Algorithms 1st Edition, Kindle Edition,” Cambridge University Press, May 2014.

[38] D. Nguyen, C. Nguyen, Thuan Duong-Ba, H. Nguyen, A. Nguyen and T. Tran, “Joint network coding and machine learning for error-prone wireless broadcast,” 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1-7, Jan. 2017.

[39] Gunst, R., Mason, R., “Regression Analysis and its Application,” Boca Raton: CRC Press, 1980.

[40] T. M. Cover and P. E. Hart, “nearest neighbor pattern classification,” IEEE Transaction on Information Theory, vol. 13, pp. 21–27, Jan. 1967.

[41] P. Hart, “The condensed nearest neighbor rule,” IEEE Transactions on

Information Theory, vol. 14, pp. 515–516, 1968.

[42] Mohammad Norouzi, David J. Fleet, Russ R. Salakhutdinov, “Hamming Distance Metric Learning,” Advances in Neural Information Processing Systems(NIPS), pp. 1061–1069, Dec. 2012.

[43] Thomas Dietterich , Eun Bae Kong, “Machine Learning Bias, Statistical Bias, and Statistical Variance of Decision Tree Algorithms”, Technical report, Department of Computer Science, Oregon State University, 1995.

[44] Segal, Mark R, “Machine Learning Benchmarks and Random Forest Regression,” Technical Report, Center for Bioinformatics & Molecular Biostatistics, University of California, San Francisco, Mar. 2004.

[45] Cortes, Corinna; Vapnik, Vladimir N., “Support-vector networks,” Machine Learning, vol. 20, pp. 273-297, Sep. 1995.

[46] https://en.wikipedia.org/wiki/Markov_chain.

[47] Eui-joong, Kim, “Introduction to artificial intelligence, machine learning, and deep learning learning through algorithms,” wikibook.co.kr, pp. 172-180, 2016.

[48] Trivedi, Kishor S, “Probability and Statistics with Reliability, Queueing, and Computer Science Applications,” John Wiley & Sons, pp. 351-419, May 2002.

[49] https://scikit-learn.org/stable/supervised_learning.html.

[50] Aurelien Geron, “Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems,” O'Reilly Media, Apr. 2017.

[51] T. Hashimoto, T. Kubo, and Y. Kitatsuji, “Lightweight paging based on location update intervals in LTE network,” 2016 IEEE International Conference on Communications (ICC), pp. 1-6, Jul. 2016.