In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 13, No.4, October-December 2025, Serial Number 52

Research Institute for Information and Communication Technology
Iranian Association of Information and Communication Technology
Affiliated to: Academic Center for Education, Culture and Research (ACECR)

Print ISSN: 2322-1437
Online ISSN: 2345-2773
Publication License: 91/13216
Editorial Office Address: No.5, Saeedi Alley, Kalej Intersection., Enghelab Ave., Tehran, Iran,
P.O.Box: 13145-799 Tel: (+9821) 88930150 Fax: (+9821) 88930157
E-mail: info@jist.ir , infojist@gmail.com
URL: jist.acecr.org

Indexed by:
- SCOPUS www. Scopus.com
- Islamic World Science Citation Center (ISC) www.isc.gov.ir
- Directory of open Access Journals (DOAJ) www.Doaj.org
- Scientific Information Database (SID) www.sid.ir
- Regional Information Center for Science and Technology (RICeST) www.ricest.ac.ir
- Magiran www.magiran.com

Publisher:
Iranian Academic Center for Education, Culture and Research **(ACECR)**

This Journal is published under scientific support of
Advanced Information Systems (AIS) Research Group and
Telecommunication Research Group, ICTRC

Acknowledgement

JIST Editorial-Board would like to gratefully appreciate the following distinguished referees for spending their valuable time and expertise in reviewing the manuscripts and their constructive suggestions, which had a great impact on the enhancement of this issue of the JIST Journal.

(A-Z)

- Afsharirad, Majid, Kharazmi University,Tehran, Iran
- Alaeiyan, Mohammad Hadi, K.N. Toosi University of Technology, Tehran, Iran
- Asghari, Seyed Amir, Kharazmi University, Tehran, Iran
- Azarkasb, Seyed Omid, K.N. Toosi University of Technology, Tehran, Iran
- Azadimotlagh, Mehdi, Persian Gulf University, Bushehr, Iran
- Behmanesh, Ali, Iran University of Medical Sciences, Tehran, Iran.
- Borna, Keyvan, Kharazmi University,Tehran, Iran
- Diaz, Raymundo, Monterrey Institute of Technology and Higher Education, Mexico
- Dobhi Givi, Sima, University of Mohaghegh Ardabili, Ardabil, Iran
- D V, Ashoka, J.S.S. Academy of Technical Education, Bengaluru, India
- Ebadati, Omid Mahdi, Kharazmi University,Tehran, Iran
- Farsi, Hassan, University of Birjand, South Khorasan, Iran
- Farsijani, Hassan, Shahid Beheshti University, Tehran, Iran
- Fazeli Veisari, Elham, Islamic Azad University, Chalus Branch, Iran
- Forouzesh, Moslem, Trbiat Modares University, Tehran, Iran
- Ghasemzadeh, Mohammad, Yazd University, Yazd, Iran
- Hejazinia, Roya, Allameh Tabataba'i University, Tehran, Iran
- Kazerouni, Morteza, Malek-Ashtar University of Technology, Tehran, Iran
- Khazaei, Mehdi, Kermanshah University of Technology, Kermanshah, Iran
- Kheirkhah, Fatemeh, ACECR, Tehran, Iran
- Kuchaki Rafsanjani, Marjan, Shahid Bahonar University, Kerman, Iran
- Maesoumi, Mohsen, Jahrom University, Shiraz, Iran
- Molazadeh,Amir Hosein, K. N. Toosi University of Technology, Tehran, Iran
- Mohajer, Amir, ICT Research Institute, Tehran, Iran
- Moayedi, Fatemeh, University of Larestan Higher Education Complex, Fars, Iran
- Moradi, Rasoul, K. N. Toosi University of Technology, Tehran, Iran
- Nangir, Mahdi, University of Tabriz, Tabriz, Iran
- Pashaeian, Matin, Amirkabir University, Tehran, Iran
- Rastegar, Abbasali, Semnan University, Semnan, Iran
- Rezakhani, Afshin, Ayatollah Ozma Borujerdi University, Lorestan, Iran
- Riahi, Noushin, Alzahra University, Tehran, Iran
- Salunke, Bharti, Poornima University, Jaipur, Rajasthan, India

- Saraeian, Shideh, Islamic Azad University, Gorgan Branch, Iran
- Soleimanian Gharehchopogh, Farhad, Islamic Azad University Urmia, Iran
- Shirmarz, Alireza, Al Taha University, Tehran, Iran
- Sharma, Divya, Chitkara University, Punjab, India
- Sharma, Deepti, Chandigarh University, Chandigarh, India
- Stanford Mphahlele, Ngoanamosadi , Tshwane University of Technology, South Africa
- Tayefeh Mahmoodi, Maryam, Research Institute for Information and Communication Technology, Tehran, Iran
- Taghavifard, Mohammad Taghi, Allameh Tabataba'i University, Tehran, Iran
- Tanhaei, Mohammad, Ilam University, Ilam , Iran
- Torabi Jahromi, Amin, Nanyang Technological University, Singapore
- Yaghoobi, Kaebeh, Ale Taha Institute of Higher Education, Tehran, Iran
- Zahedi, Mohammad Hadi, K. N. Toosi University of Technology, Tehran, Iran

Table of Contents

Journal of Information Systems and Telecommunication
Vol.13, No.4, October-December 2025, 266-277
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Adaptive PID and Fuzzy Logic Control for Yaw Attitude in LEO Satellites

Stanley E. Ajagba[1], Udora N. Nwawelu[1], Bonaventure O. Ekengwu[1*], Nnaemeka C. Asiegbu[1], Dumtochukwu O. Oyeka[1], Ifeanyi M. Chinaeke-Ogbuka[1]

[1]. Department of Electronic and Computer Engineering, University of Nigeria, Nsukka, Enugu State, Nigeria

Received: 21 Oct 2024/ Revised: 04 Sep 2025/ Accepted: 29 Oct 2025

Abstract

The significance of an effective satellite attitude control system lies in its ability to ensure that data acquisition by a Low Earth Orbit (LEO) satellite is of good quality and reliable. In this paper, the design of an adaptive Proportional Integral Derivative (PID) controller and its modified form (PIDD), which includes an additional derivative component for a microsatellite y-axis attitude control system (ACS), is presented. Additionally, a Fuzzy Logic Controller (FLC) and its enhanced version, called Adjustable Gain Enhanced FLC (AGE-FLC), were designed. Models of the amplifier, actuator, and satellite structure were developed to derive the transfer function of the LEO satellite's yaw-axis attitude dynamics. Model Reference Adaptive Control (MRAC) based Proportional Integral Derivative (PID), referred to as MRAC-PID and its modified form, MRAC-PIDD, were designed. The models of the various control systems were developed in MATLAB and were used to simulate the designed control systems. The simulation results and analysis revealed that the MRAC-PID controller offered the most efficient performance in terms of fast response and transient time, with a rise time of 1.74 seconds and a settling time of 6.19 seconds. Also, the MRAC-PIDD and AGE-FLC exhibited no overshoot, indicating efficient performance in terms of stability and smoothness in torque control. All proposed control systems for the LEO satellite yaw-axis ACS met the performance criteria, except for the PID and FLC controllers, which yielded overshoots of 12% and 21.97%, respectively. Generally, it suffices to say that the introduction of the designed adaptive PID/PIDD controllers and the AGE-FLC enhanced the system performance.

Keywords: Adaptive PID; Attitude Control System; Fuzzy Logic Controller; LEO Satellite; Yaw-axis.

1- Introduction

Satellite attitude refers to the orientation of a satellite in space, taking into account various coordinate systems [1]. The importance of satellite attitude control is evident in various areas, extending beyond communication, navigation, and earth observation. For example, in a communication satellite, it ensures that the satellite antennas are aligned with the Earth or other satellites, enabling reliable communication. Also, it maintains optimal signal strength by keeping the antenna pointed in the correct direction. The satellite's orientation in space is controlled by satellite attitude control (SAC), which ensures proper control manoeuvring. However, the flight attitude of the satellite changes to different degrees during the on-orbit flight of a satellite because of external disturbances and gravitational perturbations [2]. These disturbances acting on the satellite can cause it to shift over time, and the effect can manifest as angular variations in pitch, yaw, and roll [3]. Given that a satellite is exposed to varying disturbances, maintaining a preset attitude and a specified attitude is crucial to achieving the desired function and performance criteria [4]. Hence, for a satellite to accomplish its tasks, it is crucial to examine its control subsystem and then select the control technique that ensures the achievement of attitude adjustment and stability by improving both transient and dynamic characteristics, as well as steady-state performance [4]. Since a satellite is subject to various disturbances in orbit, its attitude and its reliability regarding data acquisition are largely dependent on the effectiveness of the SAC system. Consequently, several techniques are presented in previous studies for accurate satellite operation control. Classical and linear control schemes such as Proportional Integral and Derivative (PID) control algorithms, are prevalent implemented methods. As an example, a PID controller was used to stabilize the yaw-axis of a microsatellite by minimizing the Integral Time Absolute Error (ITAE) criterion in [5]. Similarly, other approaches involving the use of PID and its enhanced approaches in achieving

✉ **Bonaventure O. Ekengwu**
bekengwu@yahoo.com

attitude control regarding satellite yaw-axis are well documented. The PID controller uses three simultaneously coordinated computational operations to carry out corrective commands that put the plant or process response in a new state [6]. In the control of plants or processes in industries, PID controllers have been the most applied technique among several other control strategies because of their design and implementation simplicity [7].

Despite this advantage, the performance of the PID controller is largely affected by a mismatch in system parameters [8] and associated overshoot, which has led to many approaches, including intelligent algorithms, being used to tune its parameters [9]. Additionally, the PID controller is classified as a linear control system, exhibiting poor anti-interference capabilities and a significant disadvantage in its reliance on manual parameter adjustments. Therefore, several other control techniques are being implemented to address the shortcomings of the classical PID technique.

A Model Reference Adaptive Control-based PID (MRAC-PID) controller was applied in the yaw-axis stabilization of a microsatellite to enhance the settling time in [4]. PID and fuzzy-PID models were used in the SAC system. The PID offered faster steady-state, though there was certain torque oscillation, while the fuzzy-PID provided smoother and improved stability in transient and steady-state performance response [2]. Portella et al. [10] used four control moment gyroscopes (CMGs) pyramidal model to investigate the performance of a circular on-orbit flight satellite. The four CMGs' pyramidal arrangement employs either a linear quadratic tracker (LQT) with an integrator or exponential mapping control (EMC). The results indicated that the LQT has high settling time due to the use of only an integral control algorithm without proportional and derivative schemes. At the same time, the EMC showed faster but more oscillatory performance. PID, adaptive PID, and Fuzzy Logic Control (FLC) were separately compared in a laboratory nanosatellite and its testing system in [11]. The FLC rather than PID yielded significant improvement in energy consumption, convergence time, and robustness following changes in environmental conditions, which were the performance criteria, including steady state error (accuracy). Narkiewicz et al. [12] applied a PID controller whose gains are selectable for a nanosatellite attitude control and stabilization system with a generic model. In using radiation pressure of sum from solar panels with dual-mode Model Predictive Control (MPC), three-axis stabilization control of a spacecraft attitude that is under-actuated with two reaction wheels was achieved in [13]. Using a variable structured PID controller, [14] achieved attitude control of satellites by integrating a conventional PID model, trajectory planning, variable structure, and fault tolerance. The controller was designed to improve the convergence rate of the system. Enejor et al. [15] carried out a performance comparison of the PID control system and Linear Quadratic Regulator (LQR) regarding LEO satellite on-orbit flight stabilization. The study revealed that after 500 seconds, the PID was not able to stabilize the system, contrary to the LQR, which achieved the specifications for the yaw, roll, and pitch-axis. A genetic algorithm (GA) optimized fuzzy logic control system was used for attitude control in a nanosatellite by DelCastañedo et al. [16]. Considering the possible modes along the whole satellite mission, a multi-objective function cost was to optimize the fuzzy controller. Both mono and multi-objective optimizations were carried out. The system performance with mono objective optimization resulted in output that cannot be applied in practice as a result of the enormous cost of electrical power. The multi-objective optimization offered results that permit some rapid flexibility in changing the controller, including at low cost.

From the literature, it was observed that the adaptive PID controller effectively eliminates the oscillatory actions of PID, which were previously caused by initial overshoot. Consequently, the convergence time of the PID was significantly minimized for non-satellite ACS [11]. The high overshoot effect of a PID causes instability in the system's performance and impacts the smoothness of the control process. This underscores the need for a control system that will eliminate high overshoot in the system to ensure smooth and improved stability during the on-orbit flight operation of the satellite. Therefore, taking into account the advantage of an adaptive PID controller, an adaptive control system with a modified PID model (called MRAC-PIDD) and an enhanced FLC algorithm (called AGE-FLC) was proposed in this paper for the LEO satellite yaw-axis ACS. The main objective is to enhance the dynamic response of LEO satellite yaw-axis attitude. The specific objectives are to determine the dynamic equations of the components of satellite yaw-axis ACS, develop control systems based on the algorithms of the proposed solutions, and evaluate the performance of different control systems, including the proposed solution, via simulation tests in MATLAB/Simulink.

2- System Design

The yaw-axis attitude control system is modelled as a closed-loop control system, as shown in Figure 1. The system comprises a controller, an amplifier, a Direct Current (DC) motor, a satellite system, and a feedback sensor with unity gain. The controller regulates a system's behaviour by adjusting inputs to achieve the desired output. Its functions include monitoring the system's state,

comparing the actual output to the desired input, calculating the error, and sending a control signal to adjust the system's input. An amplifier increases the amplitude of a signal. In a feedback control system, it amplifies feedback signals to improve system stability and accuracy. DC motor is an electric motor that converts direct current electrical energy into mechanical energy. In control system, it can be used for position control, speed control, or motion control. In the control system, the feedback sensor measures physical parameters such as position and provides feedback to the controller, thus allowing it to adjust the system's input; and helps detect deviation from the desired set points.

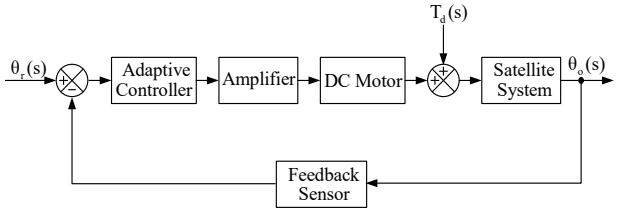

Fig. 1 Closed Loop Network of Satellite Yaw-axis ACS

For the yaw-axis ACS shown in Figure 1, the objective of the design is to ensure that the yaw-axis attitude or angle is stabilized by providing a suitable control manoeuvre that returns and keeps the satellite on its referenced or target attitude. In the figure, $\theta_r(s)$ is the reference or target attitude, while $\theta_o(s)$ is the actual attitude. Hence, to achieve stabilization and effective control of yaw-axis attitude at any instant, $\theta_r(s) = \theta_o(s)$. To meet the design objective, the proposed system is expected to achieve the following design criteria for a typical LEO satellite system: overshoot of ≤ 5%, settling time of ≤ 10 seconds, and zero steady-state error [1],[15].

2-1- Mathematical Modelling

For the closed-loop control system shown in Figure 1, the closed-loop model for the yaw-axis ACS, neglecting the controller, consists of the amplifier, DC motor, and satellite structure components. The mathematical models of these components, which comprise the amplifier, DC motor, and satellite structure, are derived subsequently.

2-1-1- Mathematical Model of Amplifier

The dynamic equation of the amplifier with gain k_a is defined in terms of the output voltage by [1]:

$$V_a(s) = K_a V_i(s) \tag{1}$$

where $V_a(s)$ is the amplifier voltage, K_a is the amplifier gain, and $V_i(s)$ is the input voltage.

Hence, the open-loop gain of the amplifier can be expressed by Eq. (2).

$$K_a = \frac{V_a(s)}{V_i(s)} \tag{2}$$

2-1-2- Mathematical Model of DC Motor

An armature-controlled DC motor is schematically represented in Figure 2. In the figure, DC motor is shown to have armature resistance and inductance R_a and L_a respectively, input or armature voltage, V_a , armature current I_a , and motor back electromotive force (EMF) of V_b that make up the electrical component of the motor. The mechanical components are the motor moment of inertia J_a , damping ratio of the motor B_a, and the motor shaft angular position $\theta(t)$.

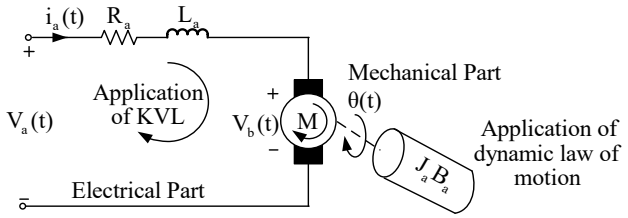

Fig. 2 Schematic Diagram or Armature Controlled DC Motor

Kirchhoff's voltage law (KVL) and dynamic law of motion application in the electrical and mechanical parts of the DC motor results to Eq. (3), Eq. (4), and Eq. (5) as seen in [17],[18].

$$V_a(t) = R_a I_a(t) + L_a \frac{dI_a(t)}{dt} + V_b(t) \tag{3}$$

$$V_b(t) = K_b \omega_m(t) = K_b \frac{d\theta(t)}{dt} \tag{4}$$

$$T_m = K_m I_a \tag{5}$$

where ω_m is the motor angular speed and is equal to the derivation of the motor angular position or displacement, K_b is the motor back EMF constant, T_m is the motor toque, and K_m is the torque constant of the motor. Eq. (5) can further be expressed by Eq. (6).

$$T_m = J_a \frac{d^2\theta(t)}{dt} + B_a \frac{d\theta(t)}{dt} \tag{6}$$

By substituting Eq. (4) into Eq. (3), Eq. (5), and Eq. (6), Eq. (7) and (8) are established.

$$V_a(t) = R_a I_a(t) + L_a \frac{dI_a(t)}{dt} + K_b \frac{d\theta(t)}{dt} \quad (7)$$

$$J_a \frac{d^2\theta(t)}{dt} + B_a \frac{d\theta(t)}{dt} = K_m I_a \quad (8)$$

Taking the Laplace transform of Eq. (7) and Eq. (8), and assuming zero initial conditions gives Eq. (9) and Eq. (10).

$$V_a(s) = L_a s I_a(s) + R_a I_a(s) + K_b s\theta(s) \quad (9)$$

$$J_a s^2\theta(s) + B_a s\theta(s) = K_m I_a(s) \quad (10)$$

By equating Eq. (9) and Eq. (10), the armature current was eliminated leading to the formulation presented in Eq. (11).

$$\frac{V_a(s) - K_b s\theta(s)}{sL_a + R_a} = \frac{J_a s^2\theta(s) + B_a s\theta(s)}{K_m} \quad (11)$$

The constants K_b and K_m are usually given as $K_b = K_m = K$ in most DC motor [1]. Therefore, Eq. (11) can be expressed in terms of transfer function as the ratio of the motor angular position to the armature input voltage as presented in Eq. (12).

$$\frac{\theta(s)}{V_a(s)} = \frac{K}{s[(J_a s + B_a)(L_a s + R_a) + K^2]} \quad (12)$$

2-1-3- Mathematical Model of Satellite System

The load torque (T_L) due to the torque delivered by the DC motor (T_m) and the disturbance torque (T_d) as shown in Figure 1 is given by Eq. (13).

$$T_L = T_m + T_d \quad (13)$$

The moment of inertia J of the entire system consists of the motor moment of inertia J_a and the moment of inertia of the satellite structure or body J_l about axis of rotation at the center of mass [1]. Given the associated vicious friction B of the satellite structure (i.e. the load) and its actual angular position $\theta_a(t)$ about the yaw-axis, the load (satellite) torque assuming $T_d = 0$ is given by Eq. (14):

$$T_L = T_m = J \frac{d^2\theta_o(t)}{dt} + B \frac{d\theta_o(t)}{dt} \quad (14)$$

The Laplace transform of Eq. (14) assuming zero initial condition, is given by Eq. (15).

$$T_m(s) = Js^2\theta_0(s) + Bs\theta_0(s) \quad (15)$$

The transfer function describing the satellite's body dynamics, specifically the relationship between the actual angular position or attitude of the yaw-axis and the input motor torque $T_m(s)$, is provided in Eq. (16).

$$\frac{\theta_o(s)}{T_m(s)} = \frac{1}{s(Js + B)} \quad (16)$$

Table 1 shows the description of the values of the physical parameters for amplifier, DC motor, and satellite structure of Low Earth Satellite (LEO).

Table 1: Parameters of the Yaw-axis ACS [1]

Definition	Symbol	Value
Amplifier	K_a	10
Motor a constant	K	0.01 Nm/A
Resistance of motor	R_a	1 Ω
Inductance of motor	L_a	0.5 H
Damping ratio of motor	B_a	0.01 Kgm^2
Moment of inertia of motor	J_a	0.1 Nms
Moment of inertia of satellite	J	2.5 Kg m^2
Damping ratio of satellite	B	1.17 Nms

Substituting the values for the parameters in Table 1 into Eq. (2), Eq. (12), and Eq. (16) yields the numerical expressions for amplifier transfer function gain, the DC motor transfer function, and the satellite body transfer function as presented in Eq. (17), Eq. (18), and Eq. (19), respectively.

$$\frac{V_a(s)}{V_i(s)} = 10 \quad (17)$$

$$\frac{\theta(s)}{V_a(s)} = \frac{0.01}{0.05s^3 + 0.105s^2 + 0.0101s} = \frac{0.2}{s^3 + 2.1s^2 + 0.202s} \quad (18)$$

$$\frac{\theta_o(s)}{T_m(s)} = \frac{1}{2.5s^2 + 1.17s} \quad (19)$$

The yaw-axis ACS is represented with a block diagram in terms of the transfer function of the amplifier, DC motor, satellite body, and unity gain feedback sensor, assuming zero torque disturbance, in Figure 3.

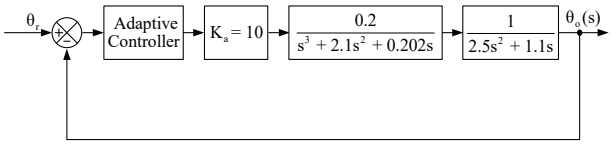

Fig. 3 Closed Loop Network of Satellite Yaw-axis ACS with Zero Torque Disturbance

2-2- Design of MRAC Based PIDD Controller

The yaw-axis ACS for LEO satellite is achieved by designing an MRAC based PIDD controller. In designing MRAC, many approaches such as Massachusetts Institute of Technology (MIT) rule, augmented error theory, and Lyapunov theory can be used. Nevertheless, approach based on MIT is used in this work. In designing a MRAC, it is required that the error and cost function be determined as shown as part of this subsection. Figure 4 is the proposed MRAC based PIDD controller for yaw-axis attitude determination.

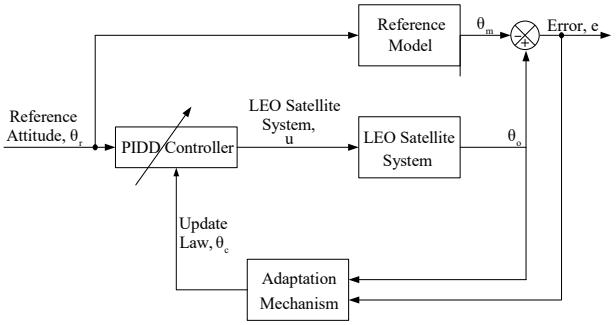

Fig. 4 The Proposed MRAC-PIDD Control System for LEO Satellite Yaw-axis ACS

2-2-1- Update Law Mechanism

The update law or the adjustment mechanism is achieved from the following mathematical expressions defined as follows.

In starting the development of the adaptation mechanism that governs the update law, the deviation (or error) due to the difference of the plant response θ_o and the output $\theta_{\mathrm{mod}\,el}$ of the reference model is expressed as in Eq. (20).

$$\text{Error, } e = \theta_o - \theta_m \tag{20}$$

The cost function $J_g(\theta_c)$ is defined in Eq. (21) in terms of *e* as;

$$J_g(\theta_c) = \frac{1}{2}e^2 \tag{21}$$

The cost function is minimized such that θ_c can be sustained in negative gradient direction of J_g defined by Eq. (22).

$$\frac{d\theta_c}{dt} = -\gamma \frac{\partial J}{\partial \theta_c} = -\gamma e \frac{\partial e}{\partial \theta_c} \tag{22}$$

The change in θ_c is established by Eq. (22) with respect to time so as the cost function can be minimized to zero. Also, $\partial e/\partial \theta_c$ is regarded as the sensitivity derivative [19]. It shows the error change with respect to the gain, which is a quantity of positive value for the controller's adaptation mechanism [20],[21]. A reference model, whose performance characteristic the satellite positioning system is to follow, is established next, and this forms the design objective of the MRAC.

Assuming the system transfer function is defined as $KG_p(s)$ where K is a quantity of unknown value and $G_p(s)$ represents transfer function of the plant. Let an expression be defined for the reference model as presented in Eq. (23).

$$G_m(s) = K_o G_p(s) \tag{23}$$

where K_o is a quantity of known value. Eq. (20) can be redefined resulting to Eq. (24).

$$E(s) = KG_p(s)U(s) - K_o G_p(s)U_c(s) \tag{24}$$

where $KG_p(s)U(s) = \theta_o$, $U(s)$ is the control input to the plant, $K_o G_p(s)U_c(s) = \theta_m$ and $U_c(s)$ the reference model input.

Thus, Eq. (25) defines the control law:

$$U(s) = \theta_c \times U_c(s) \tag{25}$$

By substituting Eq. (24) into Eq. (23), a partial derivative is applied with the resulting expression defined by Eq. (26)

$$\frac{\partial E(s)}{\partial \theta_c} = KG_p(s)U_c(s) = \frac{K}{K_o}\theta_m \tag{26}$$

Equating Eq. (22) and Eq. (26) gives Eq. (27)

$$\frac{d\theta_c}{dt} = -\gamma e \frac{K}{K_o}\theta_m = -\gamma^1 e \theta_m \tag{27}$$

where $\gamma^1 = \gamma K/K_o$ and the update law is represented by Eq. (27).

2-2-2- Determination of the Reference Model

For an MRAC, it is usually required to define a reference model $G_m(s)$. Since the satellite structure is a second-order model, in this work, the performance of the entire model of the satellite system is constrained to that of a second-order reference model dynamic and steady-state that will meet the stated performance specifications or criteria for the yaw-axis ACS. Hence, Eq. (28) defines the reference model:

$$G_m(s) = \frac{\omega_n^2}{s^2 + 2\zeta\varpi_n s + \omega_n^2} \quad (28)$$

where ω_n is the system natural frequency and ζ is the system damping ratio. The determination of the values for these quantities is carried out with Eq. (29).

$$O_p = e^{-\pi\zeta/\sqrt{1-\zeta^2}} \quad (29)$$

where O_p is the peak percentage overshoot of value 2%. Solving Eq. (29) gives Eq. (30).

$$\log_e\left(\frac{5}{100}\right) = -\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\log_e^e \quad (30)$$

This results in $\zeta = 0.78$. With the value of the damping ratio determined, the natural frequency of the system is determined using Eq. (31).

$$T_s = \frac{4}{\zeta\omega_n} \quad (31)$$

Thus $\omega_n = 5.13\ \text{rads}^{-1}$, Substituting this value into Eq. (28) gives Eq. (32).

$$G_m(s) = \frac{26.32}{s^2 + 8s + 26.32} \quad (32)$$

Eq. (32) is the designed referenced model.

2-2-3- Design of PID and PIDD Controllers

Figure 5 shows a simplified structure of PID control system used to achieve three-term process control.

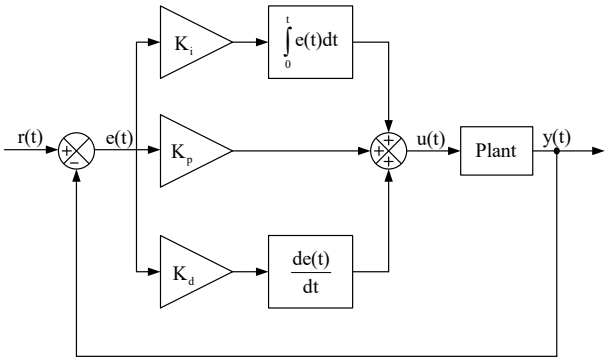

Fig. 5 PID Control System Representation

The mathematical expression of PID controller can be determined by analyzing Figure 5. Hence, $r(t)$, $e(t)$, and $u(t)$ represents the desired input, error and the control command. Furthermore, K_p, K_i, K_d are the gains of the PID controller: proportional gain, integral gain, and derivative gain. The output $y(t)$ is related to $r(t)$ by Eq. (33).

$$e(t) = r(t) - y(t) \quad (33)$$

The proportional, integral and derivative computation carried on the error as it is fed into the PID controller results to a control action given by Eq. (34).

$$u(t) = K_p e(t) + K_i\int_0^t e(t)dt + K_d\frac{de(t)}{dt} \quad (34)$$

Eq. (34) is a PID control variable in time domain. Thus, the Laplace transform of PID control variable assuming zero initial condition is given by Eq. (35).

$$U(s) = K_p E(s) + K_i\frac{1}{s}E(s) + K_d sE(s) \quad (35)$$

Or in a more simplified form as presented in Eq. (36):

$$C(s) = K_p + K_i\frac{1}{s} + K_d s \quad (36)$$

where, $C(s) = U(s)/E(s)$ and is the PID controller. The gains of the PID controller obtained by tuning the

MATLAB/Simulink PID block are [specific gains]. Thus, the designed PID controller is given by Eq. (37).

$$C(s) = 1.98 + \frac{0.0215}{s} + 1.85s \quad (37)$$

PID controllers typically introduce overshoot in control systems, which is addressed by adding an extra *D* element in this work. Hence, PIDD is a modified PID controller with extra *D* element and it is given by [22]:

$$C(s) = K_p + K_i\frac{1}{s} + K_{d1}s + K_{d2}s^2 \quad (38)$$

where $C(s) = U(s)/E(s)$ and is the PID controller. The gains of the PID controller obtained by tuning the MATLAB/Simulink PID block are [specific gains]. Therefore, the designed PIDD controller is given by Eq. (39):

$$C(s) = 1.98 + \frac{0.0115}{s} + 1.85s + 1.85s^2 \quad (39)$$

2-3- Design of Fuzzy Logic Controller

In designing an FLC, the components involved are fuzzification, defuzzification, rule base, and interference mechanism. Decision making is performed by the interference mechanism. Figure 6 is a block diagram of fuzzy logic control of satellite yaw-axis ACS.

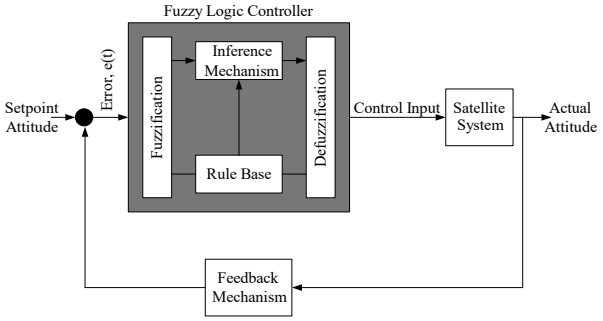

Fig. 6 Fuzzy Control System Representation

Fuzzification is performed, which involves transforming a crisp fuzzy input set into linguistic variables. The input sets or variables in this case are the error E and the change in error (ΔE). Proper scaling factors were used to scale the error and change in error for the yaw-axis attitude. The resulting linguistic variables from the fuzzification are negative big (NB), negative medium, negative small (NS), zero (ZO), positive small (PS), positive medium (PM), and positive big (PB). The corresponding fuzzy logic control rule table is shown in Table 2. For the designed FLC, each input has 3 membership functions (MFs) while the output has 5 MFs.

Table 2: Rule Base Table of the FLC

E/ΔE	NE	ZO	PO
NE	NB	NM	ZO
ZO	NM	ZO	PM
PO	ZO	PM	PB

The designed FLC was realized using the Mamdani model in MATLAB/Simulink environment. Centroid was used for the purpose of defuzzification. The inputs and outputs were modelled using the triangular MF. The resulting shapes of the triangular MFs are shown in Figure 7. The Simulink model for the designed FLC control system is shown in Figure 8. The output of the developed fuzzy model was enhanced by an adjustable gain to improve its performance as in [6]. This is called Adjustable Gain Enhanced FLC (AGE-FLC).

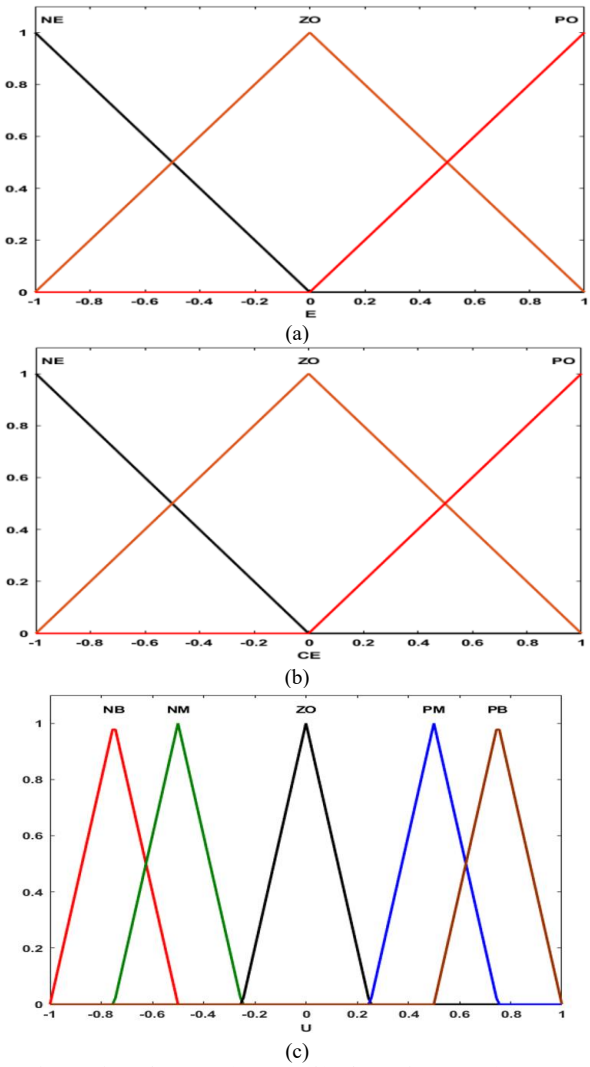

Fig. 7 Triangular MFs: (a) Error (b) Change in Error (c) Output

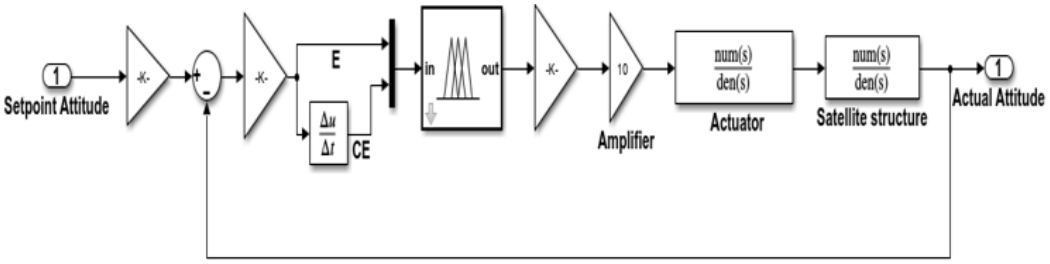

Fig. 8 Simulink Model of the Fuzzy Logic Controlled ACS

3- Results and Discussion

3-1- Analysis of the System without Controller

In this scenario, simulation analysis was conducted to investigate the performance of the microsatellite yaw axis attitude in the absence of a controller. That is no controller was introduced as a subsystem in the attitude control system (ACS) so as to ensure the stabilization of the satellite yaw angle and tracked the desired yaw-axis attitude while ensuring that the system performance criteria that include rapid convergence (that is reaching steady state as fast as possible, which is defined by the settling time in second) with little or no cycling (defined in terms of peak overshoot in percentage) are met. The resulting step response of the uncompensated satellite yaw-axis ACS (Sys1) is shown in Figure 9. The numerical analysis of the step response curve is shown in Table 3.

Table 3: Time Domain Characteristics of System Without Controller

Step Response Parameter	Value
Rise time	2.17 s
Transient time	22.14 s
Settling time	22.14 s
Peak overshoot	38.66%
Final value	0.89
Steady state error	0.11

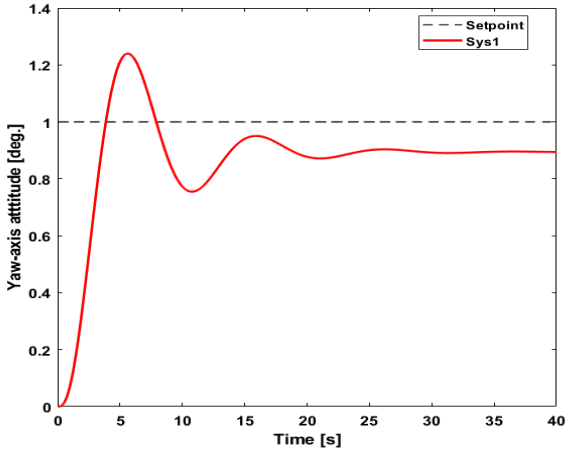

Fig. 9 Step Response of System without a Controller

Considering the step response shown in Figure 9, the numerical analysis as shown in Table 3 revealed that in the absence of a control algorithm, the system has a transient and steady-state that is characterized by rise time of 2.17 seconds, transient time and settling time 22.14 seconds respectively, peak overshoot of 38.66%, final value of 0.89 degree, and steady state error of 0.11. As shown in Figure 9, the curve reveals that in the absence of a controller, the system fails to achieve the desired attitude and suffers from high instability, which can be attributed to the magnitude or size of the overshoot. Therefore, there is a need to design a controller for on-orbit flight performance improvement in terms of yaw angle stability with significantly reduced overshoot or zero overshoot.

3-2- Analysis of PID/PIDD Control System

The simulation analysis of the PID and the PIDD controllers applied to the LEO satellite yaw axis attitude control system is presented in this section. Figure 10 shows the step response curves of the PID and the PIDD controllers. Table 4 shows the numerical analysis of the performance of the control system scenario considered using PID or PIDD controllers.

Table 4: Time Domain Characteristics of PID/PIDD

Step Response Parameter	PID	PIDD
Rise time (s)	1.93	4.26
Transient time (s)	7.97	8.34
Settling time (s)	7.97	8.34
Peak overshoot (%)	12.00	0.00
Final value	1.00	1.00
Steady state error	0.00	0.00

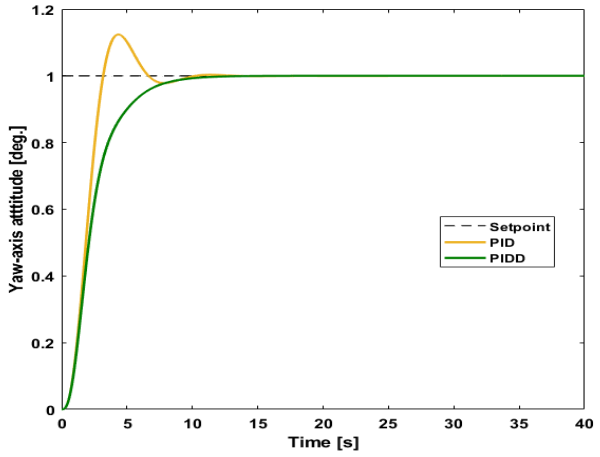

Fig. 10 Step Response of PID/PIDD Control System

Looking at Figure 10 and Table 4, it can be deduced that the PID control system showed better performance in terms of rise time and settling time than the PIDD control system. However, in terms of overshoot (or stability performance), the PIDD controller outperformed the PID controller. Though the PID control system showed good performance in terms of rise time and settling time, it did not meet all the performance criteria required of the control system, specifically the overshoot, which is 12% (i.e. > 5%). On the other hand, the PIDD control system meets the designed requirement for both settling time and overshoot: 8.34 seconds (i.e. < 10 seconds) and 0.01% (i.e. < 5%).

3-3- Analysis of the MRAC-PID/MRAC-PIDD Control System

The performances of the adaptive PID and the adaptive PIDD control systems used for the control of the yaw-axis attitude of the LEO satellite are presented in this subsection. The step response curves and the table of the numerical values obtained from the simulation analysis conducted in MATLAB/Simulink environment with respect to the designed MRAC-PID and MRAC-PIDD yaw-axis ACS for LEO satellite are presented in Figure 11 and Table 5.

Table 5: Time Domain Characteristics of MRAC Based Control System

Step Response Parameter	MRAC-PID	MRAC-PIDD
Rise time (s)	1.74	4.28
Transient time (s)	6.19	8.95
Settling time (s)	6.19	8.95
Peak overshoot (%)	3.94	0.00
Final value	1.00	1.00
Steady state error	0.00	0.00

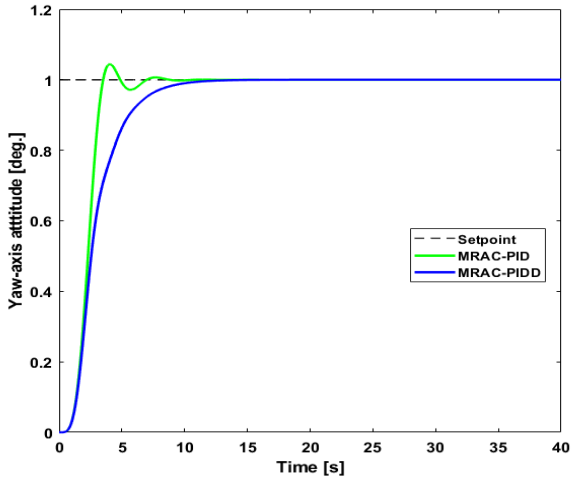

Fig. 11 Step response of MRAC-PID/MRAC-PIDD Control System

The step response curves in Figure 11 revealed that the MRAC-PID still exhibits slight oscillation though the design or performance criteria was achieved by both control systems. As shown in Table 5, the MRAC-PID yielded faster response and better convergence time in terms of rise time and settling time than the MRAC-PIDD. However, the MRAC-PIDD showed more smooth and stable performance than the MRAC-PID, which is an indication of better control torque during the operation of the satellite [2].

3-4- Analysis of Fuzzy Logic Control System

The performances of the developed FLC and its enhanced type (AGE-FLC) are presented in Figure 12 and Table 6. The simulation curves in Figure 12 reveal the step response performance of the FLC and the AGE-FLC with the gain K varied between $0.70 \leq K \leq 0.90$ for optimal response efficiency.

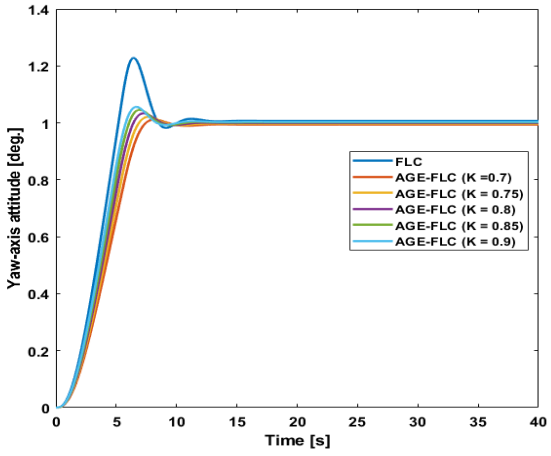

Fig. 12: Step Response of FLC/AGE-FLC Control System

Table 6: Time Domain Characteristics of FLC Based Control System

Step Response Parameter	FLC	AGE-FLC (K = 0.7)	AGE-FLC (K = 0.75)	AGE-FLC (K = 0.8)	AGE-FLC (K = 0.85)	AGE-FLC (K = 0.9)
Rise time (s)	3.33	4.50	4.23	4.26	3.81	3.64
Transient time (s)	9.52	7.01	8.18	8.32	7.87	7.67
Settling time (s)	9.52	7.01	8.18	8.32	7.87	7.67
Peak overshoot (%)	21.97	1.73	2.51	0.00	4.43	5.23
Final value	1.00	1.00	1.00	1.00	1.00	1.00
Steady state error	0.00	0.00	0.00	0.00	0.00	0.00

As shown in Figure 12 and Table 6, the FLC showed the best performance in terms of rise time, but give the worst performance with respect to settling time and overshoot. Among the AGE-FLCs, it can be seen that all met the performance criteria stated for the LEO satellite except when the gain was equal to 0.9. Thus, for optimal performance using the developed AGE-FLC for the LEO satellite yaw-axis ACS, the adjustable gain should be tuned between 0.7 and 0.85. Since stability is of utmost priority for orbiting satellites in space, the best performance is offered by AGE-FLC when K = 0.8 because it offers an overshoot of zero. It also offered smoother and more stable torque control performance [2]. Hence, amongst the AGE-FLCs, the method for K = 0.8 was used for comparison with other control systems implemented for the LEO satellite yaw-axis ACS.

3-5- Performance Comparison of Control Systems

In this section, the various control schemes implemented were compared as by the step response of the yaw-axis attitude in degree shown in Figure 13.

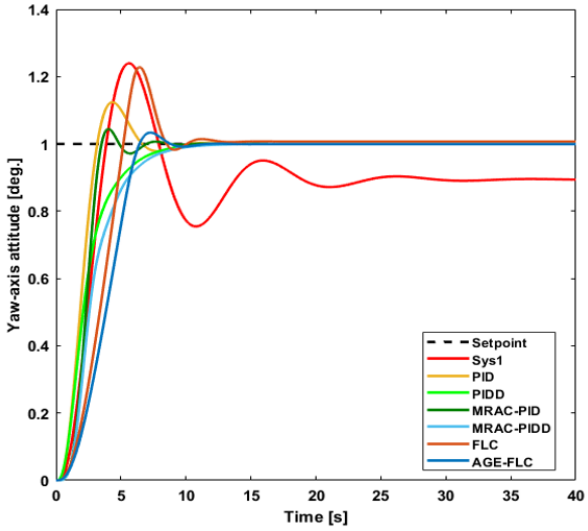

Fig. 13 Step response comparison of control systems

Among the control systems, the FLC shows the worst performance in terms of settling time and overshoot. Also, of all the control systems, only the PID and the FLC controllers did not meet the stated performance criteria considering their overshoots. The MRAC-PID controller offered the fastest response (in terms of rise time) and fastest transient time (settling time) compared to other control strategies. Thus, MRAC-PID provides the shortest time reach stability compare to other controllers. However, the MRAC-PID has an overshoot associated with its response. On the other hand, using the PIDD, MRAC-PIDD, or the AGE-FLC (K = 0.8) control system revealed that each of them effectively prevents oscillation or fluctuation of the control torque, considering the no overshoot they offered [2]. Hence, by this performance, the PIDD, MRAC-PIDD, and the AGE-FLC control systems provided the best performance in smoothness and stability.

Now, considering the settling times and overshoots of the PIDD, MRAC-PIDD, and the AGE-FLC with respect to the performance criteria established, it suffices to say that AGE-FLC provided the best performance by providing the fastest transition to stable state (steady-state) with no fluctuation in torque control handling. Generally, in terms of overall performance for meeting the stated design or performance criteria: overshoot of less than or equal to 5% and settling time of less than or equal to 10 seconds, the adaptive PID controller outperformed both the conventional PID and FLC controllers. This agrees with the experimental observation wherein adaptive PID offered the most performance compared to conventional PID and FLC for nanosatellite ACS in [11].

4- Conclusions

In this paper, adaptive controllers (MRAC-PID and MRAC-PIDD) and AGE-FLC have been developed for LEO satellite yaw-axis attitude control system (ACS). For the objective of the study to be achieved, the dynamic equations describing the yaw-axis attitude of a LEO satellite were derived. The dynamic equations were then modelled using Simulink embedded blocks in MATLAB. Each of the model components shows different features of Simulink, including transfer function block, gain block,

PID block, and fuzzy logic block. MRAC-based PID/PIDD controllers and an AGE-FLC were developed. The designed control systems were modelled and simulated in the MATLAB/Simulink environment. The results from the simulation revealed that the proposed controllers met the performance specifications of the LEO satellite yaw-axis ACS, given as overshoot of $\leq 5\%$ and settling time of ≤ 10 seconds and zero steady-state error. Notably, the MRAC-PIDD and the AGE-FLC provided negligible overshoot. This indicated that stabilization was effectively achieved and both controllers offered smoothness in control torque during the satellite operation. Generally, considering the time domain characteristic of the system response without a controller, it suffices to say that the introduction of the designed adaptive PID/PIDD controllers and the AGE-FLC largely enhanced the system performance by offering settling time less than 10 seconds and overshoot very much less than 5%. Thus, the proposed controllers met the design criteria for a typical LEO satellite system. In future work, it is recommended to integrate the PID algorithm with the FLC. Other intelligent control techniques, such as swarm algorithms or machine learning models, can be implemented with PID to reduce the settling time further. Other control system scenarios should be studied regarding the evaluation of the control systems with disturbance torque.

References

[1] C. C. Mbaocha, C. U. Eze, I. A. Ezenugu, and J. Onwumere, "Satellite Model for Yaw-axis Determination and Control using PID Compensator," International Journal of Scientific & Engineering Research, Vol. 7, No. 7, 2016, pp. 1623-1629.

[2] Y. Shan, L. Xia, and S. Li, "Design and Simulation of Satellite Attitude Control Algorithm Based on PID," Journal of Physics: Conference Series, Vol. 2355, No. 012035, 2022. pp. 1-9.

[3] H. Travis, "Introduction to Satellite Attitude Control," In Advances in Spacecraft Attitude Control, IntechOpen. 2020.

[4] P. C. Eze, and I. A. Ezenugu, "Microsatellite Yaw-axis Attitude Control System using Model Reference Adaptive Control Based PID Controller, "International Journal of Electrical and Control Engineering Research, Vol. 4, No. 2, 2024, pp. 8-16.

[5] A. T. Ajiboye, J. O. Popoola, O. Oniyide, and S. L. Ayinia, "PID Controller for Microsatellite Yaw-axis Attitude Control System Using ITAE Method," TELKOMNIKA Telecommunication, Computing, Electronics and Control, Vol. 18, No. 2, 2020, pp. 1001 – 1011.

[6] P. C. Eze, B. O. Ekengwu, N. C. Asiegbu, and T. I. Ozue, "Adjustable Gain Enhanced Fuzzy Logic Controller for Optimal Wheel Slip Ratio Tracking in Hard Braking Control System," Advances in Electrical and Electronic Engineering, Vol. 19, No. 3, 2021, pp. 231 – 242.

[7] P. C. Eze, A. E. Jonathan, B. C. Agwah, and E. A. Okoronkwo, "Improving the Performance Response of Mobile Satellite Dish Antenna Network within Nigeria," Journal of Electrical Engineering, Electronics, Control and Computer Science, Vol. 6, No. 21, 2020, pp. 25-30.

[8] B. C. Agwah, and P. C. Eze, "An Intelligent Controller Augmented with Variable Zero Lag Compensation for Antilock Braking System," International Journal of Mechanical and Mechatronics Engineering, Vol. 6, No. 11, 2022, pp. 303-210.

[9] P. C. Eze, J. K. Obichere, E. S. Mbonu, and O. J. Ononjo, "Positioning Control of Satellite Antenna for High Speed Response Performance," IPTEK, The Journal of Engineering, Vol. 10, No. 2, 2024, pp. 119-136.

[10] K. M. Portella, W. N. Schinestzki, R. M. Sehnem, L. B. da Luz, L. Q. Mantovani, R. R. Sacco, S. S. Kraemer, and P. Paglione, "Satellite Attitude Control Using Control Moment Gyroscopes," Journal of Aerospace Technology and Management, São José dos Campos, Vol. 12, 2020, pp. 94-105.

[11] A. Bello, K. S. Olfe, J. Rodríguez, J. M. Ezquerro, and V. Lapuerta, "Experimental Verification and Comparison of Fuzzy and PID Controllers for Attitude Control of Nanosatellites," Advances in Space Research, Vol. 71, 2023, pp. 3613-3630.

[12] J. Narkiewicz, M. Sochacki, and B. Zakrzewski, "Generic model of a satellite attitude control system. International Journal of Aerospace Engineering," Vol. 2020, pp. 1-17.

[13] L. Jin, and Y. Li, "Model Predictive Control-based Attitude Control of Under Actuated spacecraft Using Solar Radiation Pressure," Aerospace, Vol. 9, 2022, pp. 1-20.

[14] Y. Qi, H. Jing, and X. Wu, "Variable structure PID Controller for Satellite Attitude Control Considering Actuator Failure," Applied Sciences, Vol. 12, 2022, pp. 1-19.

[15] E. U. Enejor, F. M. Dahunsi, K. F. Akingbade, and I. O. Nelson, "Low Earth Orbit Satellite Attitude Stabilization Using Linear Quadratic Regulator," European Journal of Electrical and Computer Science, Vol. 7, No. 3, 2023, pp. 17 – 29.

[16] Á. del Castañedo, D. Calvo, Á. Bello, and M. V. Lapuerta, "Optimization of Fuzzy Attitude Control for Nanosatellites," In: K. Arai, S. Kapoor, R. Bhatia (eds) Intelligent Systems and Applications. InelliSys 2018. Advances in Intelligent Systems and Computing, Vol 869. Springer, Cham.

[17] P. C. Eze, C. A. Ugo, and D. S. Inaibo, "Positioning Control of DC Servomotor-based Antenna using PID Tuned Compensator," Journal of Engineering Sciences, Vol. 8, No. 1, 2021, pp. E9-E16.

[18] I. O. Akwukwaegbu, O. C. Nosiri, M. Olubiwe, C. F. Paulinus-Nwammuo, and E. Okonkwo, "Design of Model Following Control Integrating PID Controller for DC Servomotor-based Antenna Positioning System," SSRG International Journal of Electrical and Electronics Engineering, Vol. 10, No. 6, 2023, pp. 33-42.

[19] P. C. Eze, D. O. Njoku, O. C. Nwokonkwo, C. G. Onukwugha, J. N. Odii, and J. E. Jibiri, "Wheel Slip Equilibrium Point Model Reference Adaptive Control Based PID Controller for Antilock Braking System: A

New Approach," International Journal of automotive and mechanical Engineering, Vol. 21, No 3, 2024, pp. 11581-11595.

[20] P. C. Eze, C. A. Ugoh, C. P. Ezeabasili, B. O. Ekengwu, and L. E. Aghoghovbia, "Servor Position Control in Hard Disk Drive of a Computer Using MRAC Integrating PID Algorithm," American Journal of Science, Engineering and Technology, Vol. 2, No. 4, 2017, pp. 97-105.

[21] A. Daiifarshchi, and S. Bargandan, “Design of a Model Reference Adaptive Controller Using Modified MIT Rule for a Second Order System,” Journal of Artificial Intelligence in Electrical Engineering, Vol. 7, No. 25, 2018, pp. 7-14.

[22] M. A. Fawwaz, K. Bingi, R. Ibrahim, P. A. M. Devan, and B. R. Prusty, "Design of PIDD Controller for Robust Performance of Process Plants," Algorithms, Vol. 16, 2023, pp. 437

Journal of Information Systems and Telecommunication
Vol.13, No.4, October-December 2025, 278-288
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Automatic Concept Extraction from Persian News Text Based On Deep Learning

ZahraSadat Hosseini[1], Sayed Gholam Hassan Tabatabaei[1*]

[1].Department of Electrical and Computer Engineering, Malek-Ashtar University of Technology, Tehran, Iran

Received: 18 Dec 2024/ Revised: 04 Oct 2025/ Accepted: 02 Nov 2025

Abstract

One of the most critical issues in natural-language understanding is extracting concepts from the text. The concept expresses essential information from the text. Concept Extraction to the process of extracting and generating keyphrases that may exist or not in the text. Automatic concept extraction from the Persian news text is a challenging problem due to the complexity of the Persian language. In this paper, we first review traditional and deep learning-based models in keyphrase extraction and generation. Then, an automated Persian news concept extraction algorithm is presented, which exploits encoder-decoder models. Specifically, our proposed models use the output vector of BERT-Base and ParsBERT language models as a word embedding. The evaluation results have shown that changing the word embedding layer has improved recall, precision, and F1 measures about 3.15%. Since encoder-decoder models get inputs consecutively, the training time increases. Also, if the sentence is long, they cannot store much information from the sentences. Therefore, for the first time, we have used mT5-Base with Transformer architecture, which receives and processes data parallelly. Recall, precision, and F1 measures used for the concept extraction results of the mT5-Base model are 55.66%, 55.47%, and 55.48%, respectively. The F1 score has increased by 19.8% compared to the previous models. Therefore, this model is effective for extracting the concept of Persian news texts.

Keywords: Concept Extraction; Deep Learning; Keyphrase; BERT-BASE; ParsBERT; mT5.

1- Introduction

Many texts on different topics are published on social media every day. With the increasing volume of documents and texts, fast and reliable methods are needed to extract useful information from this vast amount of unstructured data. Concept extraction is a tool for generating and extracting keyphrases from an unstructured text that provides summary information about the text. Digital information management uses concepts for document clustering, information retrieval [1], and text summarization [2]. The concept consists of Keyphrases that may be directly present or not in the text [3]. Keyphrases can be single-word or multi-words expressions that summarize the main semantic meaning of unstructured text data and are divided into two categories [4]: absent and present Keyphrases. Unlike present Keyphrases, absent Keyphrases do not exist in the text and are implicitly mentioned in the text. In order to identify the present keyphrase in the text, keyphrase extraction algorithms are used. On the other hand, the keyphrase generation process performs the task of extracting explicit and implicit keyphrases from the text. The concept extraction is the task to extract and generate keyphrases at the same time [5].

The Internet provides people's information and contains a large amount of textual data. Therefore, it is difficult, expensive, and time-consuming for humans to extract concepts from huge documents. Hence, automatic concept extraction systems are needed [6]. So far, various automated systems have been designed for generating and extracting Keyphrases, but Persian concept extraction is still a challenge. This is for some reasons: First, most of the algorithms have been presented for the English language and little research has been done on the Persian language. Second, the structural complexity of the Persian language is higher than many languages such as English, and the other important reason is the existence of ambiguities in natural languages such as ambiguity in reference, lexical ambiguity due to polysemy, and ambiguity in distinguishing subject and object due to

✉ **Seyed Gholam Hassan Tabatabaei**
tabatabaei@mut.ac.ir

omission from the sentence. Therefore, other methods are needed to extract the concepts of Persian texts.

Given the mentioned challenges and the importance of extracting concept from Persian texts, this paper focuses on proposing an automated method for extracting concepts from Persian news texts. The proposed method is based on deep learning models, leveraging their ability to process large volumes of text data and capture complex patterns and semantic relationships. These models enable a more nuanced understanding of language, making them well-suited for tasks such as concept extraction and keyphrase generation, particularly in the context of Persian language processing.

2- Related Work

Concept extraction refers to the extraction and generation of Keyphrases. It is different from the text summarization process. Therefore, we review this literature in two parts: 1) Extracting and generating of keyphrases methods, and 2) summarization methods.

2-1- Extracting and Generating of Keyphrases

In early works, the text keyword extraction methods often included three steps: First, some keywords are extracted as text concept candidates. Second, the extracted concepts are refined using prior knowledge. As a result, the probability of reaching higher-level concepts increases, and finally, keywords are scored based on statistical information or prior knowledge [7]. Some automated keyword extraction systems are based on supervised approaches that attempt to map the sample space into two classes "key semantic units" and "non-key semantic units". Witten et al. [8] proposed a simple keyword extraction algorithm (KEA) that selects candidate keywords by calculating the TF-IDF (Term Frequency–Inverse Document Frequency) [9] and obtains the final keywords by the Naive Bayes algorithm. Zhang et al. [10] extracted Keyword from Chinese documents using a conditional random fields algorithm. The conditional random field model performed better than other machine learning methods such as linear regression and support vector machine model. Barla et al. [11] extracted key concepts instead of words for document classification using naïve Bayes model and obtained better results on news documents compared to TF-IDF keyword model.

Unlike supervised methods, unsupervised methods use unlabeled data to extract keywords. These can be divided into the graph-based, statistics-based, and language model-based methods. Khozani et al. [12] presented a statistical-based algorithm to extract keywords. In the first step, they removed the redundant words and weighed the remaining words with the TF-IDF criterion. Then using the n-gram method and based on the words' position, the weight of the words was updated and the key sentences were determined. Finally, keywords were extracted from the selected sentences. The experimental results have shown that the inference time and accuracy of this method for extracting keywords are high. Unsupervised statistical techniques such as KP-MINER [13], RAKE [14], and YAKE [15] use statistical features of texts to extract keywords. These methods are more complex due to a large number of operations. TextRank [16], SingleRank [17], and their extensions TopicRank [18] and ExpandRank [17] are graph-based methods that construct graphs to rank words based on their location in the graph. These techniques perform poorly in identifying cohesiveness between different words that constitute a keyword. Language model-based techniques utilize language model-derived statistics to extract keywords from the text [19-20]. Doostmohammadi et al. [21] conducted a comprehensive assessment to compare the performance of supervised and unsupervised methods in news keyphrase extraction and generation. Their research showed that 1) contrary to expectations, KP-Miner is better than the supervised method, 2) unsupervised approaches based on statistics are also better than graph-based methods, 3) The use of machine translation evaluation such as BLEU and ROUGE provides a more realistic evaluation for the task of keyphrase extraction and generation, 4) And all the keyphrase are not explicitly mentioned in the text. Therefore, generative models are needed to extract the non-expressed or absent keyphrases.

Recent research underscores the importance of discourse-level analysis for concept extraction, emphasizing the need to account for relations spanning sentences. For example, dependency graphs and discourse relations effectively capture linguistic structures. Techniques such as Clause Matching, as highlighted by I-Hung Hsu et al. [22], leverage dependency arc types to extract cohesive concepts from multi-sentence texts. This perspective aligns with the growing use of deep learning models, which excel in modeling complex semantic relationships and discourse structures [22].

Ontology-based concept extraction builds on these methods by integrating domain-specific knowledge to refine candidate keyphrases and associate terms with hierarchical structures. Gayathri and Kannan [23] developed a system for Ayurvedic texts that leverages domain ontologies, semantic weighting with TF-IDF, and k-Nearest Neighbors (kNN) classifiers for document classification, achieving superior results compared to traditional methods [24].

Deep learning-based methods have outperformed other machine learning methods in numerous natural language processing tasks, especially in keyphrases generation. The idea behind these methods is to learn complex features directly from data. Yuan et al. [25] adopted the RNN-based seq2seq architecture with a copy mechanism for keyphrase generation. This architecture predicts a group of keyphrases with variable length, which is considered its

advantage. Swaminathan et al. [26] proposed a CGAN generative architecture to generate keyphrases from research articles. Sun et al. [27] designed the DivGraphPointer architecture by combining traditional graph-based ranking methods and neural network-based approaches to generate keyphrases. The CopyRNN architecture was presented by Meng et al. [28] for keyphrase generation, consisting of an encoder for learning the representation of the text and a decoder for generating keyphrases based on that representation. Various modifications of the CopyRNN architecture have been proposed recently. Zhang et al. [29] proposed another architecture called CopvRNN to manage the repetition of keywords during generation based on the CopyRNN architecture. This architecture uses a bidirectional GRU for encoding and a forward GRU for decoding. CopyRNN-based architectures consistently predict N keyphrases for any input text, while in real-world examples, the number of keyphrases may vary among different texts and should be determined based on the document's content. Chen et al. [30] improved the performance of the generative model using an integrated model. The integrated model distinguishes the semantic features of present keyphrases from absent keyphrases. However, this model is not trained end-to-end and only uses a bottom shared encoder to implicitly capture the hidden semantic relationship between absent keyphrase generation and present keyphrase extraction. The first research in the field of extracting and generating keyphrases from Persian news articles was done by Doostmohammadi et al. [31]. They showed that sequence-to-sequence deep models not only perform well in keyphrase generation, but also significantly outperformed common methods such as Topic Rank, KPMiner, and KEA in keyphrase extraction.

Glazkova and Morozov [32] investigated fine-tuned generative models for keyphrase selection in Russian scientific texts, such as mT5, and mBART. Their experiments revealed that mBART achieved the best performance in in-domain evaluations, surpassing baseline methods across multiple domains such as mathematics, history, medicine, and linguistics. This study highlighted the efficacy of generative models for multilingual keyphrase extraction tasks, particularly in scientific domains. Recently, large language models (LLMs), such as GPT-4, have demonstrated impressive performance across various tasks without requiring fine-tuning. Glazkova et al. [5] also explored the use of LLMs for keyphrase generation, specifically for Russian scientific texts. Their result shows, mBART consistently outperforms LLMs and other baselines in in-domain evaluations, achieving up to higher F1 scores in fields like Mathematics and Medicine. Overall, LLMs can perform well on a variety of tasks without needing additional fine-tuning for each specific task. However, the performance of LLMs is highly reliant on the quality and design of the prompts. A poorly designed prompt may lead to inaccurate or irrelevant results, limiting the model's reliability and consistency [33].

Thomas and Vajjala [34] introduced an approach to separate present keyphrase extraction and absent keyphrase generation into distinct tasks, focusing on increasing diversity in absent keyphrase generation through specialized attention mechanisms. Their findings demonstrated improved performance across six English datasets, particularly for absent keyphrase generation tasks, emphasizing the role of distinct modeling strategies for present and absent keyphrases.

A recent study by Song et al. [35] explores the use of prompt-based unsupervised keyphrase extraction by leveraging large pre-trained language models like T5. The authors demonstrate that designing effective prompts significantly impacts performance, with complex prompts performing better for long documents. However, simple prompts often suffice for shorter texts. Their experiments on six benchmark datasets, including Inspec, SemEval2010, DUC2001, SemEval2017, Nus, and Krapivin, which are primarily English datasets, reveal that well-crafted prompts can significantly enhance keyphrase extraction performance, and automating prompt generation could further improve efficiency and scalability in real-world applications [35]. Similarly, Shen and Le [36] proposed the TAtrans model, which leverages title attention and sequence order embeddings to enhance keyphrase generation. The model showed superior performance across several datasets, including Chinese abstracts, showcasing the potential of Transformer-based methods for keyphrase tasks in diverse languages [36].

Most of the work done on keyphrase extraction and generation has focused on non-Persian texts. Therefore, in this paper, we focus on concept extraction (keyphrase generation and extraction) from Persian news texts. Given the success of language models such as T5 in various languages, we have, for the first time, employed the mT5 language model to extract and generate keyphrases from Persian news texts.

2-2- Summarization

In the present era, alongside progress in scientific and technological fields, there is a remarkable surge in the volume of accessible data. Consequently, is beneficial to have concise information that encapsulates the essence of the original document while occupying a reduced space. Although human-generated text summarization offers advantages such as precision, comprehensiveness, and coherence, it remains a laborious and costly undertaking [37]. Summarization is the process of compressing the source text into a brief version, which contains the key information of the source text. There are two types of summarization: abstractive and extractive [38]. Extractive methods choose

essential sentences, phrases, or paragraphs from the source text to form a summary while abstractive summarization methods use linguistic methods to generate a brief text [38-39]. The abstractive summarization might contain words that are not explicitly present in the source text [39]. Most of the research has been done on extractive summarization. Recently, researchers have turned to abstractive summarization. Abstractive summarization is a complex and challenging task due to the complexities of natural language text [40]. Abstractive summarization methods are broadly divided into three categories: 1) structure-based approaches, 2) semantic-based approaches, 3) deep learning-based approaches. The structure-based approach filters the most important information from the text using abstract or cognitive algorithms and includes template-based methods, tree-based methods, and ontology-based methods. Semantic-based approaches take text as input and construct a semantic representation of it. Information item-based methods, semantic graph-based methods, and multimodal semantic models use the semantic-based approach [41].

Some abstractive summarization algorithms give more scores to the summaries with more words in common with the source text and pay less attention to the semantic similarity between generated sentences and the source text. Therefore, Salemi et al. [42-43] presented a deep learning-based architecture to extract text summaries. This architecture is a pre-trained encoder-decoder model that has shown good performance in summarizing Persian text. Similarly, Shanthakumari et al. [44] used the PEGASUS model for abstractive summarization, which generates summaries by capturing key information from the original text, offering improved coherence and relevance. Their experiments demonstrated that PEGASUS, a transformer-based model, excels at generating human-like summaries by maintaining semantic integrity and reducing redundancy, addressing some of the common limitations of previous methods. Furthermore, research [45] into clinical text summarization highlights PEGASUS's capacity to distill large textual datasets into concise, coherent summaries, demonstrating comparable advantages in the medical domain where context, precision, and relevance are crucial. Liu et al. [46] proposed a hybrid summarization approach combining fine-tuned mT5 and large language models like ChatGPT, specifically evaluated on the LCSTS dataset—a large-scale Chinese short-text summarization corpus. Their approach involved using mT5 to generate initial summaries, which were refined by ChatGPT to enhance fluency and coherence, achieving high ROUGE scores and addressing key limitations in traditional models. Notably, T5's ability to treat all NLP tasks as a text-to-text problem allows it to achieve superior performance in both semantic accuracy and context preservation. This is due to its encoder-decoder transformer architecture, where the encoder captures context from the input text and the decoder generates corresponding outputs, making it particularly effective for tasks like summarization. Additionally, T5's self-attention mechanism, a core feature of the transformer architecture, enables it to focus on the most relevant parts of the input text, improving its ability to generate coherent and contextually accurate summaries. These strengths were leveraged by Liu et al. [46], where mT5 played a critical role in generating initial summaries before refinement by ChatGPT.

Encoder-decoder-based models, including the T5 model, have demonstrated good performance in both summarization and key phrase extraction tasks. Since encoder-decoder models are specifically designed and fine-tuned for tasks such as keyphrase extraction or generation, the aim of this paper is to propose a model based on the encoder-decoder architecture for extracting and generating keyphrases from Persian text, specifically news texts. The main contributions of our paper are: 1) we modify the base Encoder-Decoder [28] to extract the Persian text concept. 2) Then we change its word embedding layer by using the BERT-base [47] and ParsBERT [48] language model and present a modification of it 3) And finally, for the first time, we use the pre-trained Multilingual T5 (mT5-Base) model [3] to Persian text concept extraction. Our proposed models obtained significant results in extracting the concept of Persian news text.

The rest of this paper is organized as follows: Section 3 describes the proposed method, then the experimental setup is described in Section 4, and the experimental results are given in Section 5. Finally, it is concluded in section 6.

3- Proposed Method

The proposed method consists of two phases: pre-processing and the extension of deep learning-based language models for concept extraction. Pre-processing converts the data into a suitable format and making the process of calculations and extraction of information faster and simpler. The output of the pre-processing step is fed into the input of deep learning-based architectures. The details of each step are as follows:

3-1- Pre-processing

We use the Perkey dataset to evaluate our proposed model. This dataset has been preprocessed, as described in [21], and is publicly available. The preprocessing includes removing sentences containing specific keywords from Persian web pages and JavaScript code. Since some texts use different encodings and languages, it is necessary to unify the text to improve its analysis. For example, two Arabic letters, "ک" and "ی" are converted into their Persian equivalents. In addition to the preprocessing performed in [21], we applied further preprocessing to normalize the data using the Hazm library. The normalization process,

carried out with Hazm, consists of seven steps, which are detailed as follows:

- Analyzing the "ء ", which is a non-vowel letter and its different spellings, and correcting them.
- Removing the mentioned letter in the first step from the end of a word (such as modifying 'شهداء' to 'شهدا').
- Removing letters consisting of ' ', ' ', ' ', and ' ', from the words.
- Converting Arabic and English numbers into Persian equivalents.
- Correcting written half-spaces.
- Removing extra spaces and half-spaces used in the text.
- Correcting two-part words which are incorrect.

After normalization, the Tokenizer, a tool of BERT-Base [37], is used to split words into tokens. Tokenization recognizes the boundaries between words in texts and assigns a specific identifier to each semantic unit. As a result, a dictionary is created to convert the input text into a sequence of numbers and identifiers. Deep learning-based models are fed with the same dimensions. Since, in this paper, the BERT-BASE language model is used for word embedding, the maximum length for each token is considered 512.

3-2- The Proposed Concept Extraction Model

In this paper, the models used to extract the concept of news texts are pre-trained mT5-Base [3] and a modified encoder-decoder model. The encoder-decoder framework tries to extract the present keyphrases from the text and predict the absent keyphrases. On the other hand, with the transfer learning technique, the knowledge obtained from pre-trained models can be generalized to solve other tasks. In the following, the structure of the base models will be explained first. Then the proposed architectures are described:

3-2-1- Encoder-Decoder Structure

The encoder-decoder model was first introduced in 2014 by Chao. et al. [49] to solve translation problems. The encoder-decoder architecture [28] consists of two streams: an encoder path containing RNN blocks to learn hierarchical features from the input text. If $x = (x_1, x_2, \dots, x_T)$ is the input sequence, the hidden representation vector $h = (h_1, h_2, \dots, h_T)$ is obtained by applying the non-linear function 'f' on the x at the time step t and the previous hidden state. Then by applying the non-linear function q on the hidden representation vector, the concept vector c is obtained according to Eq. (2):

$$h_t = f(x_t, h_{t-1}) \quad (1)$$

$$c = q(h_1, h_2, \dots, h_T) \quad (2)$$

The second stream also includes RNN blocks, and its purpose is to convert the concept vector into keyphrases. Hence, this path is called the decoder. In each time step, the non-linear function f takes the concept vector, the output of the previous hidden state, and the predicted word at the time step t-1 as input and produces the hidden state s_t.

$$s_t = f(y_t, h_{t-1}, c) \quad (3)$$

Then, using the conditional language model, the predicted word y_t is obtained [21, 28].

In general, the encoder-decoder model has worked well in solving natural language processing problems, especially the generation of keyphrases, but it also has several drawbacks: 1) It is difficult to train the model for long sentences because the information containing the relationship between words is lost as the sentence length increases. Therefore, the model's accuracy in generating the main keyphrases decreases. 2) The vocabulary words of RNN models consist of a limited number of words (e.g. 30,000 words in [49]). Therefore, some keyphrases may not be included in these vocabulary words. 3) Most language models use common methods such as Word2Vec, Elmo, etc. for embedding words, but these methods do not accurately capture the relationships between words. On the other hand, the word embedding layer is the most critical part of the concept generation algorithm because its output is used as an input for the encoder-decoder model. Therefore, the design of a strong and appropriate word embedding layer is needed.

3-2-2- The Proposed Encoder-decoder Models

Inspired by the work of Meng et al. [28], we developed an encoder-decoder model utilizing bidirectional LSTM blocks for concept extraction. The structure of the proposed model, shown in Fig. 1, incorporates BERT-BASE or ParsBERT as the embedding layer, along with attention and copy mechanisms, to enhance performance. This architecture consists of two primary components: a contextual embedding layer and a modified encoder-decoder framework. The encoder processes the input text, leveraging the contextualized embeddings provided by BERT-BASE or ParsBERT, while the decoder generates output sequences. The details of each component are:

Embedding layer: According to Fig. 1, the encoder uses BERT-BASE or ParsBERT to generate contextual embeddings for the input text. In fact, we generate word embedding for textual data using the word embedding layer of ParsBERT [48] and BERT-BASE [47] models and propose BERT-BASE+Encoder-Decoder and ParsBERT+Encoder-Decoder models for the concept extraction. To use the BERT-BASE word embedding layer, the process of fine-tuning the model should be done. In this process, a list of 512 symbols is entered into the network and a 768-dimensional vector is generated. This vector is used as the input of the encoder-decoder model. Embedding from BERT-BASE would be different for the different occurrences of a word as it generates embedding's based on the context of the sentence. Other advantages of the BERT-BASE are: First, unlike other Encoders, the

BERT-BASE Encoder receives the entire sequence of words simultaneously. As a result, this is considered a bidirectional model that can learn the relationship of a semantic unit with all surrounding units [47]. Second, since BERT-BASE receives all the words of a text at once, the Masked Language Model (MLM) technique is used to train the model. In this technique, some words are randomly masked during training to increase the model's ability to learn the concept of the input sentence.

The BERT-BASE model is considered a multilingual model because it has been trained on 104 different languages. The extension of the BERT-BASE model for the Persian language under the name Pars Bert [48] was presented by Farahani et al. This model has been trained on Persian documents from various topics (such as science, novels, and news). Our experiments have shown that changing the word embedding layer using language models, especially BERT-Base, has led to improved evaluation criteria (see Section 5).

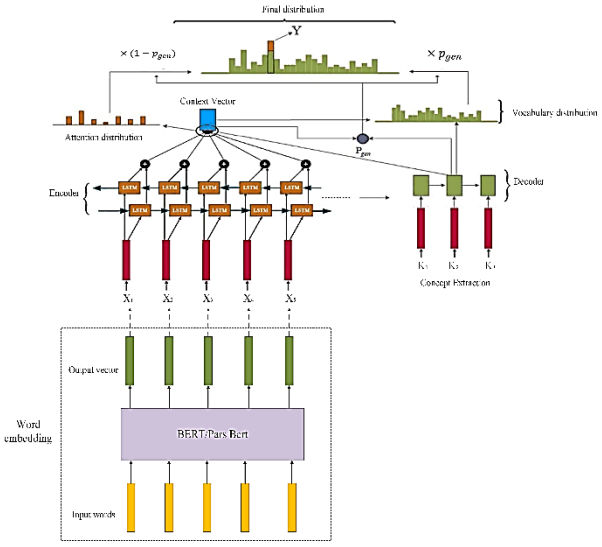

Fig. 1 The general scheme of the proposed Encoder-Decoder models with BERT-BASE /ParsBERT word embedding layer

Modified encoder-decoder: According to the explanation in Section 3.2.1 and similar to [28], the proposed architecture consists of an encoder-decoder for concept extraction, but we use bidirectional LSTMs instead of RNNs in the encoder. This is because bidirectional LSTMs, with their gating mechanisms, are better at capturing long-term dependencies and mitigating the vanishing gradient problem. These properties allow the model to effectively learn contextual relationships in both forward and backward directions, leading to more accurate concept extraction, especially in cases of complex or lengthy input texts.

Also, similar to [28], the decoder generates output by leveraging attention and copy mechanisms, addressing key challenges in sequence generation. Different words of a sentence have different importance for generating each output at each time step [50-52]. Therefore, the attention mechanism receives the output of the encoder's LSTM blocks and assigns a different weight to each of them to generate the final output. On the other hand, the copy mechanism copies certain parts of the source text exactly in the output. In this way, important key phrases that may not be present in the LSTM vocabulary are considered for concept generation.

By blending generative and extractive strategies, the final output is determined by a soft-switch parameter, p_{gen} , which dynamically adjusts the balance between generating tokens from the vocabulary and copying tokens from the input text [28].

3-3- MT5-Base Structure

The pre-trained mT5-Base model is an extension of the T5 (Text-To-Text Transfer Transformer) model which is considered an advanced version of BERT-based models. The T5 model is built using the transformers architecture, so its input and output can be text sequences. The transformer is a sequence-to-sequence model that consists of several blocks, which are connected as shown in Fig. 2: 1) an encoder block combined of a multi-head self-attention module, a position feed-forward network (FFN), residual connections to prevent gradient vanishing problem, and batch normalization layers, 2) and a decoder block, which has additional cross-attention modules between multi-head self-attention modules and position-based FFNs. The attention mechanism, as a core block of the transformer, is well-suited for long-range dependencies modeling, which is achieved by the adaptive weighting of the features according to the importance of the input. The main feature of this model is the use of relative positional embedding instead of sinusoidal positional embedding [53]. Relative positional embedding is a method for explicit and effective encoding of positional information, representing the relative position of a word in an input sentence as a vector or scalar.

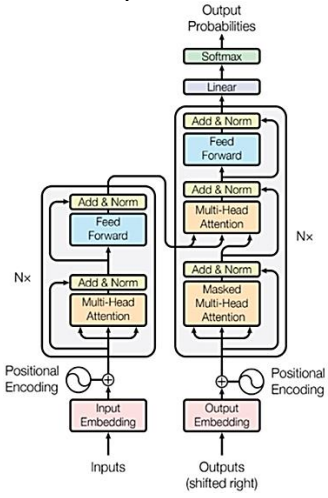

Fig. 2 The model architecture of The Transformer that was used in the mT5-Base model [47].

The use of Transformer blocks has enhanced the ability of the T5 model to perform multiple NLP tasks, including summarization, machine translation, question-answering, and classification [54]. The T5 model has been trained on a considerable amount of English texts, it cannot be generalized to other languages. The mT5-Base model is presented to solve this problem, which supports 101 different languages [55]. MT5-Base is capable of zero-shot learning and can be used in NLP tasks, including concept extraction. In this paper, we use the transfer learning technique and the pre-trained mT5-Base model to extract concepts from Persian news texts. For this purpose, we first convert the news data into text format. Then we load the mT5-Base model and use the simpleT5 class built on PyTorch-lightning and Transformers to train our model.

4- Experimental Setup

4-1- Dataset

All of the methods are validated on a subset of the Perkey dataset which includes 395,645 Persian news articles collected from 6 websites and news agencies. Each news article has at least 3 keyphrases and provides comprehensive information: {title, keyphrases, body, summary, category, URL} [21]. This dataset is divided into three subsets: training (345645 news articles), validation (2500 news articles), and test (2500 news articles) portions. The analysis carried out in [21] has shown that 31.44% of all keyphrases are not present in the text of news articles. Additionally, the number of keyphrases in news texts varies from 2 to more than 9. All this shows that the Perkey dataset can provide diverse examples with enough information to train deep learning models.

4-2- Training

The proposed method is implemented in PyTorch and evaluated on a computing server with a 3090 GPU. In the training process, the Negative Log Likelihood Loss function and Adam optimizer (initial learning rate = 10^{-4}, gradient clipping=0.1) are used.

4-3- Evaluation criteria

Various criteria were used to evaluate the performance of the proposed method in text concept extraction. First, three common criteria, namely Precision, Recall, and F1 score were used. These criteria are defined in formulas Eq. (4) to Eq. (6), respectively.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 = 2 * \frac{precision*recall}{precision+recall} \quad (6)$$

Where 'TP' is the number of true keyphrases, 'FP' is the number of false keyphrases, 'TN' is the number of true non-keyphrases, and 'FN' is the number of false non-keyphrases. Furthermore, the results of traditional models were examined in terms of ROUGE-1 and ROUGE-2 metrics. The ROUGE-1 criterion refers to the overlap of unigrams (a subsequence of n words) between the candidate summary and the reference summary. While the ROUGE-2 criterion refers to the overlap of bigrams between the candidate summary and reference summary. According to the ROUGE definition, Precision and Recall criteria are described by formulas Eq. (7) and Eq. (8), respectively.

$$Precision = \frac{number\_of\_overlapping\_words}{total\_words\_in\_system\_summary} \quad (7)$$

$$Precision = \frac{number\_of\_overlapping\_words}{total\_words\_in\_system\_summary} \quad (8)$$

5- Experimental Results and Analysis

To confirm the performance of the proposed models in extracting the concept of Persian news texts, the test results were analyzed from different perspectives:

5-1- Keyphrase Extraction

Table 1 presents the results of keyword extraction on the Perkey dataset based on ROUGE-1 and ROUGE-2 criteria for traditional models. It is observed that the KEA model performs better than other methods in terms of precision and recall. In addition, Table 2 shows the quantitative evaluation results of all methods in extracting the keyphrases of the test set from the Perkey dataset. As can be seen, the supervised learning method performs better than the statistical models and graph-based methods due to the use of labeled data. Compared to traditional methods, deep learning-based methods have achieved better results in extracting keyphrases from Persian texts due to automatic feature extraction. Therefore, the good performance of the encoder-decoder model can be seen from an increase in the F1 score to 43.04%.

Table 1: The performance of traditional models for extracting keywords on the Perkey dataset

Method		ROUGE-1			ROUGE-2		
		Precision	Recall	F1	Precision	Recall	F1
Statistical Models	TF-IDF	36.34%	27.91%	29.83%	5.51%	4.78%	4.76%
	KP-Miner	39.89%	26.06%	29.12%	5.52%	4.37%	4.47%
	YAKE	18.99%	21.81%	18.69%	3.31%	4.35%	3.40%
Graph-based Model	Single Rank	20.40%	32.48%	23.59%	4.98%	9.96%	6.19%
Supervised Model	Kea	38.14%	29.39%	31.39%	6.46%	5.88%	5.72%

Table 2: Comparison of the different keyphrase extraction methods on the Perkey dataset

Method		Dataset	Precision	Recall	F1
Statistical Models	TF-IDF	Perkey	17.24%	20.60%	18.77%
	KP-Miner	Perkey	19.00%	19.48%	19.24%
	YAKE	Perkey	7.26%	8.20%	7.70%
Graph-based Model	Single Rank	Perkey	5.32%	6.71%	5.94%
Supervised Model	Kea	Perkey	18.37%	22.26%	20.13%
Deep learning-based Models	Encoder-Decoder model	Perkey	37.24%	62.87%	43.04%
	Proposed ParsBERT+Encoder-Decoder model	Perkey	38.70%	65.01%	44.83%
	Proposed BERT-BASE +Encoder-Decoder model	Perkey	39.41%	64.68%	45.32%
	mT5-Base	Perkey	56.79%	58.54%	59.63%

On the other hand, it has been observed that by using the 768-dimensional concept vectors obtained from ParsBERT's model as the input of the proposed Encoder-Decoder model, all the evaluation criteria were improved by about two percent. Alongside this, the performance of the BERT-BASE model is better than the former because it has been trained on a large corpus of multilingual data. A high prediction F1 of 45.32% for the proposed BERT-BASE+Encoder-Decoder model confirms this. Although the performance of the proposed encoder-decoder models is superior to other methods, the precision criterion obtained from these models is significantly lower than the recall criterion. This means that the number of extracted incorrect key phrases (FP) is more than the extracted incorrect non-key phrases (FN). We applied the pre-trained mT5-Base model to overcome this problem and achieved a significant improvement (59.63% F1-score) over the results of the previous models in the keyphrase extraction task. The mT5-Base model can generate word vectors more precisely due to the use of parallel processing and relative position embedding.

5-2- Keyphrase Generation

As mentioned earlier, some keyphrases do not appear in the input text. Hence, generating absent keyphrases is a challenging task. It should be noted that traditional methods cannot generate keyphrases. Therefore, Table 3 only provides the performances of deep learning-based models for the absent keyphrases prediction task. It can be seen from Table 3 that the proposed encoder-decoder models perform better than the base encoder-decoder architecture [28] in generating absent keyphrases. Also, the findings show that using the mT5-Base model has led to the improvement of all metrics. For example, the F1 score has increased by about 25%. This is because the mT5-Base model is a multilingual model and fine-tuning it on Persian news texts helps to improve the accuracy of keyphrase prediction results. It should be noted that deep learning models must be trained on large amounts of data. Hence, Fine-tuning the pre-trained model is very useful when a small training dataset is available.

Table 3: Comparison of the different keyphrase generation models methods in the Perkey dataset

Method		Dataset	Precision	Recall	F1
Deep learning-based Models	Encoder-Decoder model	Perkey	12.38%	34.60%	17.46%
	Proposed ParsBERT+Encoder-Decoder model	Perkey	14.84%	42.01%	21.04%
	Proposed BERT-BASE +Encoder-Decoder model	Perkey	15.40%	41.93%	21.52%
	mT5-Base	Perkey	44.58%	44.39%	46.86%

5-3- Concept Extraction

The concept of a text includes both absent and present keyphrases. Table 4 presents the results related to the overall performance of all deep learning-based methods, i.e. generating absent keyphrases and extracting present keyphrases. For the proposed BERT-BASE +Encoder-Decoder model the F1-score increased by approximately 3.15%. This shows that using BERT-BASE's language model for word embedding is effective. Also, as expected, after the proposed encoder-decoder models, the best performance belongs to the mT5-Base model. The overall performance of the proposed models, i.e. BERT+Encoder-Decoder and ParsBERT+Encoder-Decoder over the entire dataset are summarized in Tables 5 and 6. It can be seen from both tables that the proposed Encoder-decoder models have predicted fewer incorrect keyphrases compared to the base encoder-decoder. Also, the keyphrases generated by the mT5-Base model are more consistent with the true keyphrases.

Table 4: .Comparison of the different Concept Extraction methods in the Perkey dataset

Method		Dataset	Precision	Recall	F1
Deep learning-based Models	Encoder-Decoder model	Perkey	31.54%	46.75%	35.68%
	Proposed ParsBERT+Encoder-Decoder model	Perkey	33.24%	49.73%	37.99%
	Proposed BERT-Base +Encoder-Decoder model	Perkey	34.23%	50.91%	38.83%
	mT5-Base	Perkey	55.47%	55.66%	55.48%

Table 5: The output of the models - Example (1).

News text	نبض «پایتخت» در دست تنابنده است هومن حاجی عبداللهی مجری، صداپیشه و بازیگری است که در همه این عرصه‌ها فعالیت دارد، اما با حضور در سریال «پایتخت» بیشتر توانایی‌هایش در زمینه بازیگری به نمایش گذاشته شد. شیرینی‌های نقش رحمت شاسی در این سریال محبوب تلویزیون مدیون بازی هوشمندانه حاجی عبداللهی است که با توجه به استقبال مخاطبان باعث پررنگ‌ترشدن حضور این بازیگر در ادامه این سریال شد. به بهانه پخش سری پنجم این مجموعه پرطرفدار پای حرف‌های این بازیگر نشسته‌ایم.
True keyphrases	تلویزیون. سریال ایرانی. بازیگران سینما و تلویزیون ایران
Encoder-Decoder model	تلویزیون. سریال ایرانی. سینما. پایتخت. هومن حاجی عبداللهی
Proposed BERT-BASE +encoder-decoder model	بازیگران سینما و تلویزیون ایران. سازمان صدا و سیما، مجری رادیو و تلویزیون. تلویزیون. سریال ایرانی. سینمای تلویزیون. مجموعه تلویزیونی پایتخت. برنامه‌های تلویزیونی. هومن حاجی عبداللهی .
Proposed ParsBERT+encoder-decoder model	بازیگران سینما و تلویزیون ایران. مجری رادیو و تلویزیون. تلویزیون. سریال ایرانی. مجموعه تلویزیونی پایتخت. برنامه‌های تلویزیونی. هومن حاجی عبداللهی.
mT5-Base	تلویزیون. سریال ایرانی. شبکه نمایش خانگی. بازیگران سینما و تلویزیون ایران

Table 6: The output of the models - Example (2).

News text	رسمی؛ جواد نکونام با تیم ملی قرارداد یکساله بست به گزارش خبرگزاری خبرآنلاین؛ پس از ابراز علاقه کارلوس کی روش سرمربی تیم ملی به حضور نکونام در کادر فنی ایران، فدراسیون فوتبال با کاپیتان سابق تیم ملی وارد مذاکره شد. البته این مذکرات زیاد به طول نکشید و سرانجام روز گذشته قرارداد جواد نکونام با تیم ملی دیروز امضا شد. شماره 6 سابق ایران روز گذشته در فدراسیون فوتبال و در حضور مهدی تاج و محمدرضا ساکت مدیر تیم‌های ملی قراردادش را امضا کرد. قرارداد او یکساله است و در پایان هر سال تمدید خواهد شد.
True keyphrases	تیم ملی فوتبال ایران. جواد نکونام.
Encoder-Decoder model	تیم ملی فوتبال ایران. فدراسیون فوتبال. کارلوس کی روش. مهدی تاج. بازی‌های انتخابی. باشگاه پرسپولیس.
Proposed BERT-Base +encoder-decoder model	تیم ملی فوتبال ایران. جواد نکونام. فدراسیون فوتبال. کارلوس کی روش. محمدرضا نکونام. مهدی نکونام. بازی‌های انتخابی جام‌جهانی. باشگاه پرسپولیس. جام جهانی.
Proposed ParsBERT+encoder-decoder model	تیم ملی فوتبال ایران. جواد نکونام. فدراسیون فوتبال. کارلوس کی روش. محمدرضا نکونام. مهدی نکونام. باشگاه پرسپولیس. جام جهانی.
mT5-Base	تیم ملی فوتبال ایران. جواد نکونام. بازی‌های انتخابی جام‌جهانی

5-4- Title and Summary Extraction

The mT5-based models have a high potential to accurately learn the relationship between the words of a text due to their use of the transformer block. The performance of mT5-Base and mT5-Small models in extracting the title and summarizing the text is reported in Table 7. The mT5-Base model has more parameters than mT5-small and is considered a stronger model. Thus it has a better performance in summary extraction. In addition, the outputs of the mT5-Base model for two examples of news texts are shown in Tables 8 and 9, respectively. The results of Tables 9 and 10 show that the text produced by this model is conceptually very similar to the true labels (i.e. True Title and True Summarization). For example, in Table 8, although the summary text produced by mT5-Base has more words (i.e. "جواد نکونام" and "با تیم ملی") than the True summary, both sentences have the same meaning. The ROUGE score is low for two sentences that have different words but the same meaning because it only computes syntactical matches. Solving this challenge requires the definition of new criteria to consider the conceptual similarity of texts, which could be a fascinating issue for future studies.

Table 7: The output of the mT5-based models in extracting the title and summary in the Perkey dataset

Method		ROUGE-1			ROUGE-2		
		Precision	Recall	F1	Precision	Recall	F1
mT5-small [47]	Title	-	-	28.5 %	-	-	12.55%
	summary	-	-	33.88%	-	-	19.17%
mT5-Base	Title	40.69%	41.76%	41.99%	23.63%	24.32%	24.30%
	summary	38.16%	39.54%	39.75%	23.34%	24.07%	24.12%

Table 8: The output of the mT5-based models in extracting the title and summary from example 1

True Title	نبض «پایتخت» در دست تنابنده است.
True Summary	هومن حاجی‌عبداللهی که حضورش در فصل‌های جدید «پایتخت» پررنگ‌تر از قبل شده است، می‌گوید این اتفاق به دلیل استقبال مخاطبان از شخصیت رحمت شاسی افتاده است.
mT5-Base (Title)	سریال «پایتخت».
mT5-Base (summary)	هومن حاجی عبداللهی که حضورش در سریال «پایتخت» پررنگ تر از قبل شده است، می‌گوید به دلیل استقبال مخاطبان از شخصیت رحمت شاسی این اتفاق افتاده است.

Table 9: The output of the mT5-based models in extracting the title and summary from example 2

True Title	رسمی؛ جواد نکونام با تیم ملی قرارداد یکساله بست.
True Summary	کاپیتان سابق تیم ملی ایران سرانجام قراردادش را امضا کرد.
mT5-Base (Title)	جواد نکونام با تیم ملی قرارداد بست.
mT5-Base (summary)	جواد نکونام کاپیتان سابق تیم ملی سرانجام با تیم ملی قرارداد بست.

6- Conclusions

In this paper, deep-learning models are used for concept extraction from Persian news texts. First, a multi-stage pre-processing technique is applied to modify the Persian text and normalize the Persian text. Then, BERT-Base +Encoder-Decoder and ParsBERT+Encoder-Decoder models are proposed to extract the concept from news text. The proposed models utilize the output vector of BERT-BASE and ParsBERT language models for word embedding. The experimental results showed that the performance of the proposed models are significantly better than previous models. One of the disadvantages of encoder-decoder-based models is the generation of many incorrect keyphrases. The pre-trained mT5-Base model performs well in title extraction and abstractive text summarization tasks. Therefore, this model was also used to extract the concept. It was observed that this model has a significant ability to predict the concept of the text.
In general, compared to traditional methods, deep learning-based models not only extract the keyphrases of the text but also generate the missing keyphrases. Future work on the concept extraction task can also extend this study to other languages.

References

[1] S. Jones and M. S. Staveley, "Phrasier: A system for interactive document retrieval using keyphrases", in Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1999, pp. 160-167.

[2] Y. Zhang, N. Zincir-Heywood, and E. Milios, "World wide web site summarization", Web intelligence and agent systems: an international journal, Vol. 2, No. 1, 2004, pp. 39-53.

[3] E. Papagiannopoulou and G. Tsoumakas, "A review of keyphrase extraction", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 10, No. 2, 2020, p. e1339.

[4] J. Chen, X. Zhang, Y. Wu, Z. Yan, and Z. Li ", Keyphrase generation with correlation constraints," arXiv preprint arXiv:1808.07185, 2018.

[5] F. Boudin, Y. Gallina, and A. Aizawa, "Keyphrase generation for scientific document retrieval", arXiv preprint arXiv:2106.14726, 2021.

[6] S. Mehrabi, S. A. Mirroshandel, and H. Ahmadifar, "DeepSumm: A Novel Deep Learning-Based Multi-Lingual Multi-Documents Summarization System", Journal of Information Systems and Telecommunication (JIST), 2019, p. 204.

[7] K. Barker and N. Cornacchia, "Using noun phrase heads to extract document keyphrases", in Advances in Artificial Intelligence: 13th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, 2000, pp. 40-52.

[8] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevill-Manning, "KEA: Practical automatic keyphrase extraction", in Proceedings of the fourth ACM conference on Digital libraries, 1999, pp. 254-255.

[9] S. N. Kim and M.-Y. Kan, "Re-examining automatic keyphrase extraction approaches in scientific articles", in Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications (MWE), 2009, pp. 9-16.

[10] C. Zhang, "Automatic keyword extraction from documents using conditional random fields", Journal of Computational Information Systems, 2008, vol. 4, no. 3, pp. 1169-1180.

[11] M. Barla and M. Bieliková, "From ambiguous words to key-concept extraction", in 24th International Workshop on Database and Expert Systems Applications, 2013, pp. 63-67: IEEE.

[12] S. M. H. Khozani and H. Bayat, "Specialization of keyword extraction approach to persian texts", in International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2011, pp. 112-116.

[13] S. R. El-Beltagy and A. Rafea, "KP-Miner: A keyphrase extraction system for English and Arabic documents", Information systems, 2009, Vol. 34, No. 1, pp. 132-144.

[14] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents", in Text Mining: Applications and Theory, 2010, pp. 1-20.

[15] R. Campos et al., "Yake! collection-independent automatic keyword extractor", in Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Vol. 40, 2018, pp. 806-810.

[16] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text", in Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pp. 404-411.

[17] X. Wan, and J. Xiao, "Single document keyphrase extraction using neighborhood knowledge", in Proceedings of the 23rd AAAI Conference on Artificial Intelligence, Vol. 8, 2008, pp. 855-860.

[18] A. Bougouin, F. Boudin, and B. Daille, "Topicrank: Graph-based topic ranking for keyphrase extraction", in Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), 2013, pp. 543-551.

[19] T. Tomokiyo and M. Hurst, "A language model approach to keyphrase extraction", in Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment, 2003, pp. 33-40.

[20] Z. Liu, X. Chen, Y. Zheng, and M. Sun, "Automatic keyphrase extraction by bridging vocabulary gap", in Proceedings of the Fifteenth Conference on Computational Natural Language Learning, 2011, pp. 135-144.

[21] E. Doostmohammadi, M. H. Bokaei, and H. Sameti, "Perkey: A persian news corpus for keyphrase extraction and generation", in 2018 9th International Symposium on Telecommunications (IST), 2018, pp. 460-465.

[22] I. Hsu, G. Xiao, N. Premkumar, and P. Nanyun", Discourse-level relation extraction via graph pooling", arXiv preprint arXiv:2101.00124 ,2021.

[23] E. Oro, R. Massimo, and S. Domenico, "Ontology-based information extraction from pdf documents with xonto", International Journal on Artificial Intelligence Tools, Vol. 18, No. 05, 2009, pp. 673-695.

[24] M. Gayathri, and R. J. Kannan, "Ontology based concept extraction and classification of ayurvedic documents", Procedia Computer Science, Vol. 172, 2020, pp. 511-516.

[25] X. Yuan, T. Wang, R. Meng, K. Thaker, P. Brusilovsky, D. He, A. Trischler, "One size does not fit all: Generating and

evaluating variable number of keyphrases", arXiv preprint arXiv:1810.05241, 2018.
[26] A. Swaminathan, R. K. Gupta, H. Zhang, D. Mahata, R. Gosangi, and R. R. Shah, "Keyphrase generation for scientific articles using gans (student abstract) ", in Proceedings of the AAAI Conference on Artificial Intelligence, 2020, Vol. 34, No. 10, pp. 13931-13932.
[27] Z. Sun, J. Tang, P. Du, Z.-H. Deng, and J.-Y. Nie, "Divgraphpointer: A graph pointer network for extracting diverse keyphrases", in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 755-764.
[28] R. Meng, S. Zhao, S. Han, D. He, P. Brusilovsky, and Y. Chi, "Deep keyphrase generation", arXiv preprint arXiv:1704.06879, 2017.
[29] Y. Zhang and W. Xiao, "Keyphrase generation based on deep seq2seq model", IEEE access, Vol. 6, 2018, pp. 46047-46057.
[30] W. Chen, H. P. Chan, P. Li, L. Bing, and I. King, "An integrated approach for keyphrase generation via exploring the power of retrieval and extraction", arXiv preprint arXiv:1904.03454, 2019.
[31] E. Doostmohammadi, M. H. Bokaei, and H. Sameti, "Persian keyphrase generation using sequence-to-sequence models", in 2019 27th Iranian Conference on Electrical Engineering (ICEE), 2019, pp. 2010-2015.
[32] A. Glazkova, and D. Morozov, "Exploring Fine-tuned Generative Models for Keyphrase Selection: A Case Study for Russian", arXiv preprint arXiv:2409.10640, 2024.
[33] A. Glazkova, D. Morozov, and T. Garipov, "Key Algorithms for Keyphrase Generation: Instruction-Based LLMs for Russian Scientific Keyphrases", arXiv preprint arXiv:2410.18040, 2024.
[34] E. Thomas, and S. Vajjala, "Improving Absent Keyphrase Generation with Diversity Heads", in Findings of the Association for Computational Linguistics: NAACL 2024, 2024, pp. 1568-1584.
[35] M. Song, Y. Feng, and L. Jing, "A Preliminary Empirical Study on Prompt-based Unsupervised Keyphrase Extraction", arXiv preprint arXiv:2405.16571, 2024.
[36] L. Shen, and X. Le, "An enhanced method on transformer-based model for one2seq keyphrase generation", Electronics, Vol. 12, No. 13, 2023, p. 2968.
[37] N. S. Shirwandkar and S. Kulkarni, "Extractive text summarization using deep learning", in 2018 fourth international conference on computing communication control and automation (ICCUBEA), 2018, pp. 1-5.
[38] M. E. Khademi, M. Fakhredanesh, and S. M. Hoseini, "Farsi conceptual text summarizer: a new model in continuous vector space", Journal of Information Systems and Telecommunication (JIST), Vol. 1, No. 25, 2019, p. 23.
[39] M. Afsharizadeh, H. Ebrahimpour-Komleh, A. Bagheri, and G. Chrupała, "A Survey on Multi-document Summarization and Domain-Oriented Approaches", Journal of Information Systems and Telecommunication (JIST), Vol. 1, No. 37, 2022, p. 68.
[40] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E.D. Trippe, J.B. Gutierrez, and K. Kochut, "Text summarization techniques: a brief survey", arXiv preprint arXiv:1707.02268, 2017.
[41] S. Gupta, and S. K. Gupta, "Abstractive summarization: An overview of the state of the art", Expert Systems with Applications, Vol. 121, 2019, pp. 49-65.
[42] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert", arXiv preprint arXiv:1904.09675, 2019.
[43] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "Pegasus: Pre-training with extracted gap-sentences for abstractive summarization", in International Conference on Machine Learning, 2020, pp. 11328-11339: PMLR.
[44] L. Shen, and X. Le, "An enhanced method on transformer-based model for one2seq keyphrase generation", Electronics, Vol. 12, No. 13 , 2023, p. 2968
[45] N. Datta, "Extractive Text Summarization of Clinical Text Using Deep Learning Models", in 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), 2024, pp. 1-6.
[46] F. Liu, C. Xiong, " A Generative Text Summarization Method Based on mT5 and Large Language Models", in 2023 Eleventh International Conference on Advanced Cloud and Big Data (CBD), 2023, pp. 174-179.
[47] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805, 2018.
[48] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "Parsbert: Transformer-based model for persian language understanding", Neural Processing Letters, Vol. 53, 2021, pp. 3831-3847.
[49] K. Cho, B. v. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation", arXiv preprint arXiv:1406.1078, 2014.
[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need. Advances in neural information processing systems", Advances in neural information processing systems, Vol. 30, 2017.
[51] D. Wang, C. Hansen, L.C. Lima, C. Hansen, M. Maistro, J.G. Simonsen, and C. Lioma, "Multi-Head Self-Attention with Role-Guided Masks", in Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28–April 1, 2021, Proceedings, Part II 43,2021, pp. 432-439.
[52] T. Xiao, Y. Li, J. Zhu, Z. Yu, and T. Liu, "Sharing attention weights for fast transformer", arXiv preprint arXiv:1906.11024, 2019.
[53] S. Yildirim and M. Asgari-Chenaghlu, "Mastering Transformers: Build state-of-the-art models from scratch with advanced natural language processing techniques", Packt Publishing Ltd, 2021.
[54] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P.J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer", The Journal of Machine Learning Research, Vol. 21, No. 1, 2020, pp. 5485-5551.
[55] L. Xue, "mT5: A massively multilingual pre-trained text-to-text transformer", arXiv preprint arXiv:2010.11934, 2020.

Journal of Information Systems and Telecommunication
Vol.13, No.4, October-December 2025, 289-299
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

A Comprehensive Framework for Enhancing Intrusion Detection Systems through Advanced Analytical Techniques

Chetan Gupta[1*], Amit Kumar[2], Neelesh Kumar Jain[3]

[1].Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, Guna, India

Received: 01 Jan 2025/ Revised: 04 Oct 2025/ Accepted: 02 Nov 2025

Abstract

Intrusion detection systems (IDS) are security technologies that monitor system activity, network traffic, and settings to detect potential threats. IDS provide proactive security management, detecting anomalies and ensuring continuous monitoring. It protects critical assets, such as sensitive data and intellectual property, from unauthorized access or data breaches, preventing downtime and disruption to business operations. In this paper we present a hybrid model based on Principal Component Analysis (PCA) and XGBoost algorithms. To show the effectiveness of the proposed system, various parameters are evaluated on the standard NSL-KDD dataset. First we trained the model using trained dataset and then evaluate the performance the model using testing dataset. In proposed work the we store the data into two-dimensional structure then we standardized and take a most significance features of the data then calculate the covariance matrix, after that calculate the eigenvalues and eigenvectors of the matrix and short in the descending order and using principal component identify the new features and remove the insignificant features. The proposed model outperforms and produces 97.76% accuracy and 94.51% precision; the recall rate is 93.44% and 93.97% F1-Score, which is much better than the previous proposed models. This hybrid approach is better to handle the categorical data and able to find the pattern well and the outcome of the model clearly shows the effectiveness of the proposed system.

Keywords: IDS; DOS; XGBOOST; PCA; HIDS; NIDS.

1- Introduction

A system that keeps an eye on network traffic for questionable behavior and sends out notifications when it finds it is known as an intrusion detection system (IDS) [1]. It is a piece of software that searches a system or network for malicious activities or policy violations [2][3][33]. Typically, a security information and event management (SIEM) system is used to gather data centrally or to alert any harmful activity or violation to an administrator [4][5]. In order to distinguish between hostile behavior and false alerts, a SIEM system combines outputs from many sources and applies alarm filtering mechanisms [6][7].
Additionally, intrusion prevention systems keep an eye on incoming network packets to look for any malicious activity. If they find any, they immediately send out warning messages. Finding intrusions seems to be the straightforward objective of intrusion detection [8][9]. The work is challenging, however, as intrusion detection systems only find indications of intrusions, either while or after they have occurred [34]. In actuality, they do not detect intrusions at all. This kind of proof is frequently called the "manifestation" of an assault [10][11]. The intrusion cannot be detected by the system if there is no manifestation, if the manifestation is incomplete or unreliable, or if it provides unreliable information [12][13]. An Intrusion Detection System (IDS) is a vital component of an organization's security infrastructure, monitoring network traffic and system activities for malicious actions or policy violations [14][15]. It provides early detection and response to threats, enhances security posture, and helps identify vulnerabilities [16]. IDS also support compliance with regulatory requirements, providing valuable logs and reports for audits [17]. It aids in incident investigation and forensic analysis, allowing organizations to trace the origins of an intrusion [18][19][20] . Figure 1 represent the network based IDS. There are various types of IDS including host-based, network-based. HIDS, which offer scalability, early detection, non-intrusion, and wide coverage. NIDS uses techniques like anomaly detection, threat categorization, signature-based detection enhancement, predictive analytics, behavioral analysis, user profiling, and insider threat identification. Wireless

✉ **Chetan Gupta**
chetangupta.gupta1@gmail.com

Intrusion Detection Systems (WIDS) monitor and analyses wireless network data to identify unauthorized access, malicious activity, or policy violations.

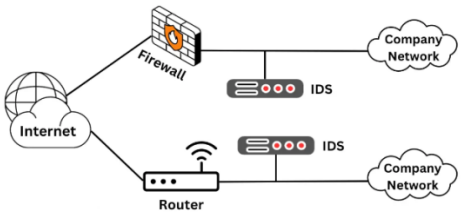

Fig. 1 Network Intrusion detection system

IDS improves network and system visibility, identifying policy violations, and supporting security policies by detecting and reporting violations [21][22]. It also encourages the implementation of best practices in network and system security, fostering a culture of proactive security management. Overall, IDS is essential for organizations seeking to enhance their security measures [23][24].

Contribution: The major contribution of the proposed hybrid model:

1. In this paper we present a hybrid model which not only able handled the numerical data but also can handle the categorical data which gives an extra benefit to use this model with any real time dataset.

2. The proposed model offer 97.76% accuracy and 94.51% precision rate that is the remarkable performance as compared to the other previous approaches.

3. In contrast to previous approaches that required modifications based upon the dataset's properties, our proposed method stands out in that we built a model that enables the use of any real-time dataset without requiring changes to the algorithm.

4. The suggested method reduces computational cost and speeds up processing by incorporating PCA to remove unnecessary and duplicate features while maintaining crucial variance.

The organizations of the study are as follows: Section 2 gives the concise overview of relevant work and the scope of improvement in the IDS. Section 3 gives the details on the problem identification or research gaps. Section 4 highlights the research objectives. Section 5 present the proposed model and step wise step explanation. In section 6 practical work and results discussion are mentioned to assess the suggested technique. Finally, Section 7 we explain the conclusion of the proposed work and the future enhancement.

2- Literature Survey

Faten Louati et al. [1] presents a novel approach to intrusion detection systems (IDS) utilizing a multi agent-based reinforcement learning architecture. The authors propose a distributed IDS framework where mobile agents monitor network activities and detect potential security breaches. This methodology enhances the scalability and flexibility of IDS, allowing for real-time threat detection and response across large network environments. The research demonstrates that mobile agents can significantly reduce the detection time and resource consumption compared to traditional IDS models.

Mahdi Soltan et al. [2] research explores the foundational principles of intrusion detection systems, highlighting the importance of anomaly detection and misuse detection techniques. The study underscores the necessity for robust IDS frameworks that can adapt to evolving cyber threats. They emphasizes the role of statistical models, expert systems, and machine learning in enhancing IDS capabilities, providing a comprehensive review of the methodologies and technologies that underpin modern IDS solutions.

Neha gupta [3] this research provides a critical analysis of the false alarm problem in intrusion detection systems. He investigates the trade-off between detection accuracy and false alarm rates, proposing several improvements to current IDS algorithms. The study highlights the need for more sophisticated data analysis techniques to distinguish between benign and malicious activities effectively. He suggests that integrating contextual information and user behavior profiling can significantly reduce false positives, thereby improving the overall efficiency of IDS.

Md. Alamin Talukder [4] examines the challenges and opportunities in network intrusion detection, focusing on the application of machine learning techniques. The authors argue that while machine learning offers significant potential for enhancing IDS, it also presents unique challenges such as training data quality, feature selection, and model interpretability. Their study provides a thorough evaluation of various machine learning algorithms and their applicability to different intrusion detection scenarios, advocating for a hybrid approach that combines machine learning with traditional detection methods.

Raghad A. AL-Syouf [5] this study offers a complete survey of anomaly-based network intrusion detection

techniques. He categorizes various anomaly detection methods, detailing their strengths and weaknesses. The authors highlight the significance of statistical, machine learning, and data mining techniques in identifying deviations from normal network behavior. The research also addresses the challenge of defining normal behavior in dynamic network environments and proposes solutions to enhance the adaptability and accuracy of anomaly-based IDS.

Vipin Kumar [6] the research investigates the use of principal component analysis (PCA) for enhancing the performance of intrusion detection systems. He demonstrates how PCA can effectively reduce the dimensionality of network data, improving the efficiency and accuracy of IDS. The study presents a detailed evaluation of PCA-based IDS models, highlighting their ability to identify designs and anomalies in large-scale network traffic. The authors conclude that PCA is a valuable tool for preprocessing data in IDS, leading to more robust and scalable intrusion detection solutions.

Mohamed H. Behiry [7] travels the intersection of machine learning and network intrusion detection, highlighting both the potential benefits and inherent challenges. The authors argue that while machine learning techniques can significantly enhance IDS performance by automating the detection of complex patterns, they also introduce issues such as the essential for high-quality training data and the difficulty of interpreting model outputs. The study proposes a hybrid approach that combines machine learning with traditional methods to balance detection accuracy and operational feasibility.

Shahad Altamimi [8] provides an in-depth review of various machine learning procedures applied to interruption detection systems. They evaluate the performance of techniques such as decision trees, support vector machines, and neural networks in detecting different types of network intrusions. The study identifies key factors influencing the effectiveness of these algorithms, including feature selection and dataset characteristics, and highlights the superiority of ensemble methods in improving detection rates and reducing false positives.

Nilesh Chothani [9] A thorough examination of deep learning techniques for network intrusion detection. The authors concentrate on using PCA and Kernel-Based Extreme Learning to identify abnormalities and categorize network traffic. According to the research, deep learning models perform better than conventional machine learning methods in terms of accuracy and flexibility against novel attack patterns, especially when they make use of hierarchical feature extraction capabilities.

Saadia Ajmal [10] this paper examines current developments in intrusion detection systems based on machine learning, with a focus on big data analytics integration. The advantages of using big data frameworks to manage the enormous volumes of network traffic data, which improves the effectiveness and scalability of IDS. The study also discusses the difficulties in processing data in real-time and the significance of choosing machine learning models that are capable of effectively analyzing and categorizing network events.

3- Problem Identification

Identifying problems in IDS is essential for improving system mechanism; here are some common challenges and issues associated with IDS:

a. The accuracy of the system in correctly identifying true positive instances among all instances is low. Hence, some irritated instances are classified.
b. The accuracy is limited for effectiveness of the IDS in a dynamic security landscape.
c. The IDS dataset is unbalanced, which results in duplicate and useless characteristics. As a result, this takes time and makes it harder to identify the assault accurately.

4- Problem Identification

a. To improve accuracy for maintaining and effectiveness of the IDS in a dynamic security landscape.
b. To improve precision for correctly identifying true positive instances among all instances.

Research Questions:

1 When compared to conventional classifiers, will hybrid PCA–XGBoost model increase intrusion detection accuracy on the NSL-KDD dataset?

2 Is it possible for XGBoost to better classify attacks by handling the complicated and unbalanced nature of network traffic data?

Hypotheses:

1 H1: Compared to solo models, the PCA–XGBoost hybrid model provides greater detection accuracy and F1-score.
2 H2: PCA improves model training speed and generalization

5- Proposed Algorithm

In proposed work we uses a hybrid technique based on principal component analysis (PCA) and XG-Boost algorithm, PCA is a unsupervised machine learning algorithm as shown in Figure 2. This is used to reduce the dimension of the dataset features and draw a pattern by reducing the variances. For experimental purpose we use the NSL-KDD dataset.

Using PCA, First we split the dataset into two part training data and testing data then we represent the data into two-dimensional structure then we standardized and take a high variance features of the data then calculate the covariance matrix, after that evaluate the eigenvalues and eigenvectors of the matrix and short in the descending order and using principal component identify the new features and remove the insignificant features as shown in algorithm 1. The equation 1 to equation 8 shows flow of PCA Algorithm. XGBoost [34] is also a machine learning algorithm which do the preprocessing of the dataset by Categorical label encoding and feature scaling by Splitting the dataset into two part usually training dataset and testing sets in 70%-30% ratio and then evaluate the model performance using detection accuracy, precision, recall and F1-score. Our result clearly shoes the improvement over past techniques. Algorithm 2 shows the step by step procedure of XGBoost algorithm. The equation 9 to equation 14 shows flow of XGBoost Algorithm

Algorithm 1: Principal Component Analysis

Step 1: Given a dataset$\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in \mathbb{R}^d$ are feature vectors and $y_i \in \{-1, 1\}$ are labels.

Step 2: Gather and prepare your dataset with features (inputs) and corresponding target variables (outputs).

Step 3: Principal Component Analysis (PCA), transform a set of possibly correlated variables (features) into a new set of orthogonal (uncorrelated) variables.

3.1 Standardization (if necessary):

If the data is not standardized (mean-centered and scaled), PCA typically begins with: Standardize

$$X: X \leftarrow \frac{X-\mu}{\sigma} \quad (1)$$

Where X is the data matrix, μ is the mean vector, and σ is the standard deviation vector across each feature.

3.2 Covariance Matrix Calculation:

Calculate the covariance matrix Σ of the standardized data X:

$$\Sigma = \frac{1}{n} X^T X \quad (2)$$

Where n is the number of data points, and XT denotes the transpose of X.

3.3 Eigen value Decomposition:

Perform eigenvalue decomposition on the covariance matrix Σ:

$$\Sigma v = \lambda v \quad (3)$$

Where:

v is the eigenvector.

λ is the corresponding eigenvalue.

λ and v satisfy the equation above.

3.4 Sorting Eigenvalues

Sort the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$, (where d is the number of original features) in descending order and arrange the corresponding eigenvectors $v_1, v_2, \ldots, v_d$ accordingly.

3.5 Choosing Principal Components

At highest to lowest, the eigenvectors are sorted according with their corresponding eigenvalues. According to the desired degree of dimensionality reduction, the individual then selects a subset among these eigenvectors to construct the newly added feature subspace. The elements that explain the most variation, or have largest eigenvalues, are often retained.

3.6 Projection onto Principal Components

Project the original data X onto the subspace spanned by the selected principal components V_k:

$$Z = XV_k \quad (4)$$

Where Z is the matrix of transformed data, where each row represents a data point projected onto the principal components.

PCA aims to maximize the variance of the projected data along the principal components, effectively reducing the dimensionality while preserving as much variance as possible.

Step 4: For each iteration t from 1 to T:

Start with an initial prediction

$$\hat{y}_i^{(0)} = 0 \quad (5)$$

Where:

$y_i$is the true label of the i-th data point.

$\hat{y}_i^{(t)}$ is the predicted label by the ensemble model after t iterations.

Step 5:

a. Compute the negative gradient of the loss function L with respect to the current ensemble's predictions:

$$g_i^{(t)} = \frac{\partial L\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}}\Big|_{\hat{y}_i^{(t-1)}=\hat{y}_i^{(t-1)}} \quad (6)$$

b. Compute the second derivative (if necessary) or other derivatives needed for the specific loss function and objectives.

c. Fit a weak learner (decision tree) to the negative gradient $g_i^{(t)}$ with certain weights (such as the learning rate η).

d. Update the ensemble model:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta . f_t(x_i) \quad (7)$$

where $f_t(x_i)$ is the prediction of the t-th tree for the i-th data point x_i.

Step 6: Regularization

Include a regularization term Ω (f_t) in the objective function to control the complexity of the ensemble:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \quad w_j^2 \tag{8}$$

Where γ and λ are regularization parameters and w_j are the weights associated with the leaves of the j-th tree.

Step 7: Final Prediction

After T iterations, the final prediction for a new data point $x_{new}$is:

$$\hat{y}_{new} = \sum_{t=1}^{T} \quad \eta . f_t(x_{new}) \tag{9}$$

This sum aggregates predictions from all trees in the ensemble, each weighted by the learning rate η.

Algorithm 2: XG-boost Algorithm**:**

Step 1: Problem Definition:

Calculate the loss function and regularized function Ω (f_t) till t^{th} iteration.

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{t=1}^{T} \Omega(f_t) \tag{10}$$

Step 2: Initialized Calculations:

Here we calculate the mean regression value.

$$\hat{y}_i^{(0)} = \text{initial guess} \tag{11}$$

Step 3: Calculate Gradients g_i and Hessians h_i:

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2} \tag{12}$$

Step 4: Gain Computation using g_i and h_i:

$$\text{Gain} = \frac{1}{2}\left[\frac{(\sum_{i\in L} g_i)^2}{\sum_{i\in L} h_i + \lambda} + \frac{(\sum_{i\in R} g_i)^2}{\sum_{i\in R} h_i + \lambda} - \frac{(\sum_{i\in L\cup R} g_i)^2}{\sum_{i\in L\cup R} h_i + \lambda}\right] - \gamma \tag{13}$$

Step 5: update predictions:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \tag{14}$$

Where $f_t\,(x_i)$ is the new tree and η is the rate of learning.

Step 6: repeat step 3 to step 5 till the final results.

Step 7: Final Calculation: It is the contributions from all trees:

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i) \tag{15}$$

The Figure 2 demonstrates the flow of hybrid PCA and XGBoost algorithm. First we trained the model using trained dataset and then evaluate the performance the model using testing dataset and the outcome of the model shows the effectiveness of the proposed approach.

By generating new synthetic samples for minority classes rather than replicating preexisting ones, SMOTE (Synthetic Minority Oversampling Technique) addresses class imbalance in the NSL-KDD dataset. By choosing a minority sample, determining nearest neighbors, and creating new data points along the lines that link them, it accomplishes this. This technique optimizes the input data, lessens over fitting, and improves the model's ability to recognize uncommon attack types.

These improved characteristics are then used by the ANN component to efficiently learn intricate non-linear attack patterns. ACO and ANN work together to improve feature quality and classification accuracy, which closes the gap between robust intrusion detection performance and feature selection efficiency.

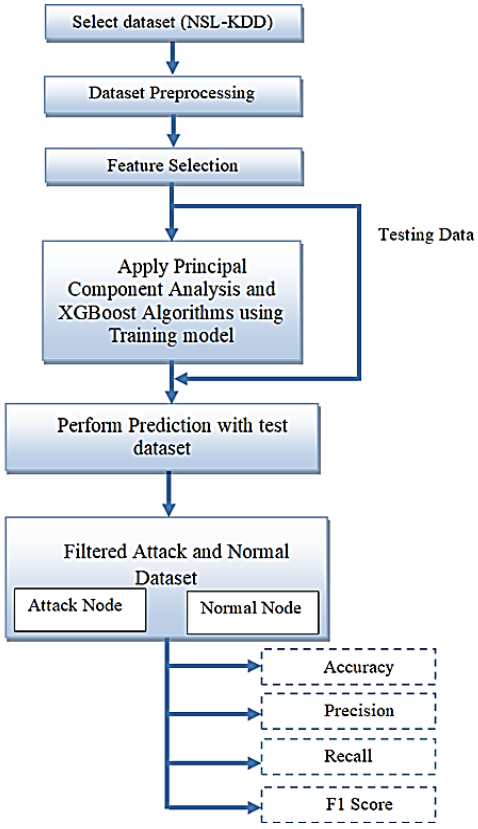

Fig. 2: Flow Chart of Proposed Methodology

6- Practical Work and Results Discussion

The following measurements are taken using Jupyter notebook and python 3.11.1 on anaconda navigator. Precision, recall, F1-Score, and accuracy are computed as follows when the suggested proposed approach is applied to concern dataset. The implementation is done using

python language and the hardware configuration we used is Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz 2.30 GHz, 8.00 GB RAM, GPU capable of 15-30 TFLOPS for deep learning. TPU capable of 90 TFLOPS for deep learning.

6-1- Description of Dataset (NSL-KDD)

A data set called NSL-KDD is proposed to address a few of the KDD'99 data set's intrinsic issues, which are listed in [25]. This dataset contains 125,973 records of a network nodes contains 23 different types of attack and a normal record [26][27]. Due to the lack of publicly available data sets for network-based intrusion detection systems, this updated version of the KDD data set still has some of the issues raised by McHugh [28] and may not be a perfect representation of current real networks. Nevertheless, we think it can still be used as a useful benchmark data set to assist researchers in comparing various intrusion detection techniques. Moreover, the NSL-KDD train and test sets have a respectable amount of records. This benefit eliminates the requirement to choose a small sample at random and makes it feasible to conduct the tests on the whole set at a reasonable cost [29]. As a consequence, assessment findings from various research projects will be uniform and equivalent [30][31][32].detected as normal. False Positives (FP): FPs is usual records inaccurately detected as anomalies. False Negatives (FN): FNs are the number of anomaly records inaccurately detected as usual.

6-2- Training of the Model:

```
import matplotlib.pyplot as plt                                   Copy  Edit

# Exact data from your screenshot
data = [
    [0, 'tcp', 'http', 'SF', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 'udp', 'private', 'SF', 44, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 'tcp', 'http', 'SF', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 'tcp', 'http', 'SF', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 'tcp', 'http', 'SF', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
]

columns = list(range(43))
df = pd.DataFrame(data, columns=columns)

# Plot and save in HD quality
fig, ax = plt.subplots(figsize=(18, 4), dpi=300)
ax.axis('off')
table = ax.table(
    cellText=df.values,
    colLabels=df.columns,
    cellLoc='center',
    loc='center'
)
table.auto_set_font_size(False)
table.set_fontsize(8)
table.scale(1.2, 1.2)

plt.tight_layout()
plt.savefig("NSL_KDD_Table_HD.png", dpi=300, bbox_inches='tight')
plt.show()
```

Fig. 3: Load dataset and identify head of IDS dataset

The figure 3 shows the different features of the imported IDS NSL-KDD dataset out of which 70% data used for the training of the model and the rest 30% used the check the performance of the trained model.

Justification for PCA and XGBoost: PCA was used to streamline datasets while preserving important patterns in order to handle the significant dimension and overlap in NSL-KDD features. Because of its flexibility, resilience, and capacity to capture intricate non-linear correlations, XGBoost was chosen as the best option for differentiating between typical and different kinds of attacks. They produce a more consistent and comprehensible intrusion detection model by bridging the gap between effective description of features and excellent detection rate.

1 Data information

Fig. 4: Pie-Chart of IDS Classes

```
[11859 rows x 166 columns]

X = scaled_data.drop(('class', 'level'], asis=1).values
y = pf('class'].values
y = n_  df('class'].values
n_components = 20)
pca = PCA(PCA.componts).fit(x)
X_reduced = pca.transform(x)
print(X_reduced)

print(['The original features before PCA: (11859,164)
print('Reduced features after PCA: (x1859,20)

[ -1.218006793e 02 -4.02536527e:01   3.11678047e 00   -- 0.570796
  -0.013000000e 00 -2.92000010e-01   1.50000000e 01    - 1.000145
  [ 1.12158735e 02 -4.15563685e:02   7.15927802e 00   -- 04415645
   -1.81899082e 01  9.02730915e-01   0.10457587e 01    - 1.792545
  [-1.21006569e 02 -4.15632351e:02   6.18706624e 00   -- 0.500483
   -1.03238605e 01  1.09063448e-01   1.70321315e 01

The original features before PCA(11859, 164)
Reduced features after PCA(11859, 20)
```

Fig. 5: PCA Evaluation of IDS dataset

In Figure 5, the feature vector is compute through PCA algorithm. Feature value computations perform based on IDS classes characteristics.

6-3- Performance Matrix Evaluation:

The performance of the model is evaluated using different standard parameter like accuracy, precision and recall rate. Table 1 shows the different outcome of the parameter to test the effectiveness of the presented model.

		Actual Value	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

Accuracy=	(TP+TN) / (TP+TN+FP+FN)
Detection Rate=	TP / (TP+FN)
Recall=	FP/(TP+TN)
F1 Score=	TP/(TP+FN)

Table 1: Estimation of Precision, Recall, F1-Score and Accuracy on Train dataset among different models and Proposed Prediction Model

Models	Precision	Recall	F1-Score	Accuracy
Linear SVC [11]	52.71	79.06	63.25	83.43
Gaussian Naïve Bayes [13]	46.65	90.00	61.45	79.63
IDS-XGbFS [30]	86.64	78.45	78.23	88.65
PCA–Firefly–XGBoost [29]	93.52	87.57	87.22	92.71
IDS-XGbFS [31]	91.87	82.14	84.75	84.45
XGBoost-PCA (Proposed)	98.37	98.60	98.48	99.45

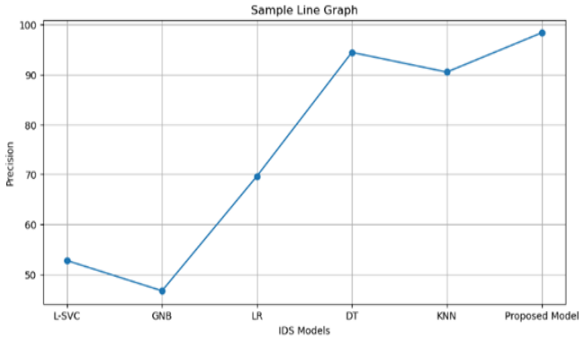

Fig. 6: Graphical Analysis of Precision among different IDS models (Train data).

The Figure 6 demonstrates that the suggested model provides superior precision when compared to other models in the context of IDS model. The proposed model perform outperforms by an improvement of 3.92% in terms of precision.

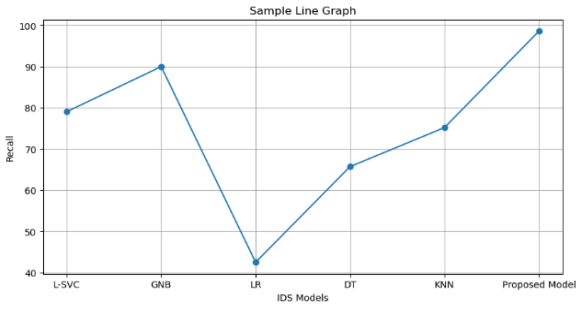

Fig. 7: Graphical Analysis of Recall among different IDS models (Train).

The figure 7, demonstrates that the suggested model provides superior recall when compared to other models in the context of IDS model. The proposed model perform outperforms by an improvement of 8.6% in terms of recall.

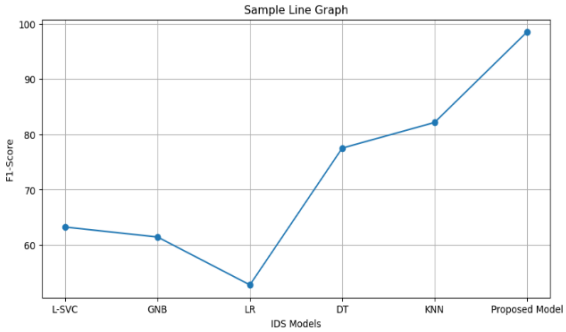

Fig. 8: Graphical Analysis of F1-Score among different IDS models (Train data).

The Figure 8 demonstrates that the suggested model provides superior F1-Score when compared to other models in the context of IDS model. The proposed model perform outperforms by an improvement of 6.33% in terms of F1-Score.

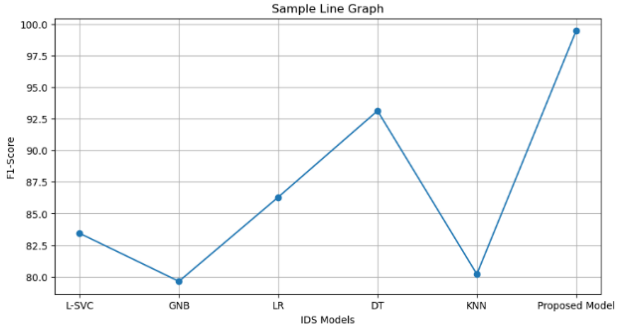

Fig. 9: Graphical Analysis of Accuracy among different IDS models (Train data).

The Figure 9 demonstrates that the suggested model provides superior Accuracy when compared to other models in the context of IDS model. The proposed model

perform outperforms by an improvement of 6.33% in terms of Accuracy.

Table 2: Estimation of Precision, Recall, F1-Score and Accuracy on Test dataset among different models and Proposed Prediction Model

Models	Precision	Recall	F1-Score	Accuracy
Linear SVC [11]	55.66	78.96	65.29	84.35
Gaussian Naïve Bayes [13]	46.75	86.20	60.62	79.11
IDS-XGbFS [30]	87.25	86.31	88.92	95.2
PCA–Firefly–XGBoost [29]	87.25	86.31	88.92	95.2
IDS-XGbFS [31]	91.33	85.49	87.20	90.42
XGBoost-PCA (Proposed)	94.51	93.44	93.97	97.76

Table 2 show the improvement in the Accuracy, precision and other parameters of the proposed model over previous models which clearly shows the effectiveness of the model. For the proposed work k-fold cross-validation is 10-fold is applied to evaluate performance stability across different data splits. A 95% confidence interval (CI) can be calculated for these metrics to measure result consistency and variation. Also the significant p-values is ($p < 0.05$).

The capacity of the XGBoost-PCA model to cover a larger variety of attack patterns results in a modest increase in false positives while significantly lowering false negatives, which accounts for the slight drop in precision (0.59%). Because ignoring an attack (false negative) usually has more serious repercussions than a false alert (false positive), this compromise is beneficial in detection of intrusions.

Therefore, the XGBoost-PCA model shows overall improved and appropriate outcomes, regardless of the somewhat lower precision, confirming its effectiveness in enhancing intrusion detection systems on the NSL-KDD dataset.

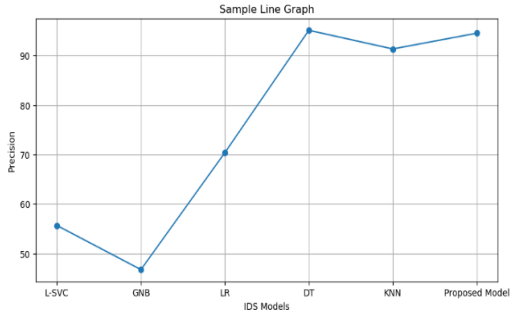

Fig. 10: Graphical Analysis of Precision among different IDS models (Test data).

As can be seen in the above Figure 10, the suggested model provides superior precision for IDS when compared to previous models. When compared to the Decision Tree, proposed model has a 0.59% decrease in precision.

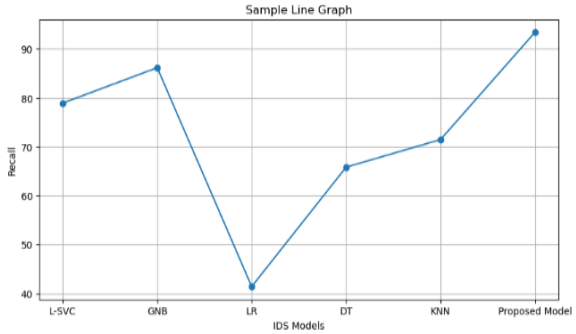

Fig. 11: Graphical Analysis of Recall among different IDS models (Test data).

As can be seen in the above Figure 11, the suggested model provides superior recall for IDS when compared to previous models. When compared to the Gaussian Naïve Bayes, proposed model has a 7.24% improvement in recall.

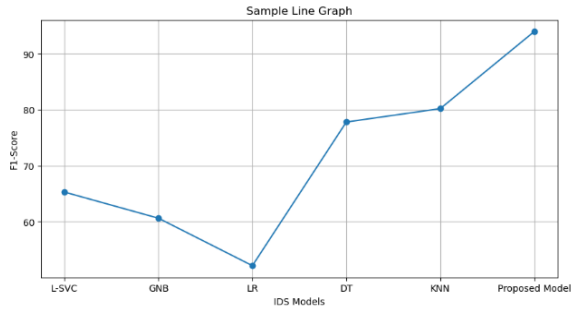

Fig. 12: Graphical Analysis of F1-Score among different IDS models (Test data).

As can be seen in the above Figure 12, the suggested model provides superior F1-Score for IDS when compared to previous models. When compared to the KNN, proposed model has a 13.77% improvement in F1-Score.

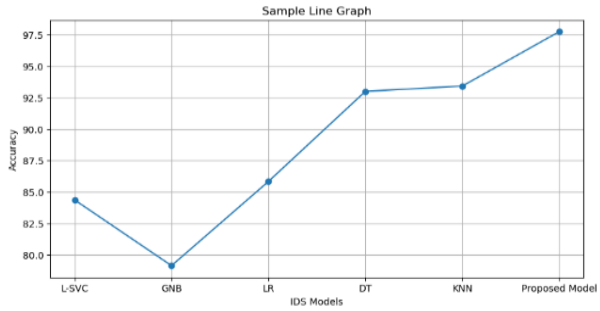

Fig. 13: Graphical Analysis of Accuracy among different IDS models (Test data).

As can be seen in the above Figure 13, the suggested model provides superior Accuracy for IDS when compared to previous models. When compared to the KNN, proposed model has a 4.34% improvement in Accuracy.

The proposed XG-Boost and PCA shows better results in terms of accuracy, precision and all other compared parameter that are evaluated as compared to other Ids algorithm including Linear SVM [11], Gaussian Naïve Bayes [13], Logistic Regression [12], and KNN [14]. Figure 14 display the validation outcome as shown in figure 14.

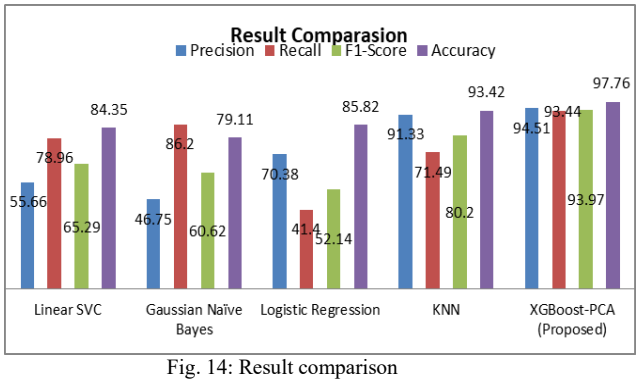

Fig. 14: Result comparison

This study has significant limitations in spite of its effectiveness. Particularly, the NSL-KDD dataset—which, despite its widespread use, does not accurately capture the intricacies of contemporary network traffic and changing attack patterns—is utilized to assess the suggested PCA-XGBoost architecture. Furthermore, low-variance characteristics that could include important information for identifying certain uncommon or complex assaults may be eliminated using PCA, which could affect the detection of minority classes. Additionally, the architecture depends on offline batch processing for training, which could need to be modified for continuous, real-time data streams on fast networks. Additionally, the interpretability of judgments after PCA transformation is still restricted, and XGBoost hyper parameter adjustment might be computationally demanding. These restrictions show how much more effort is required to expand the framework in large-scale, diverse systems.

7- Conclusions

Machine learning (ML)-based intrusion detection systems (IDS) have shown potential to revolutionize security by providing more accurate, adaptive, and scalable solutions. In this paper we present a hybrid model based on Principal Component Analysis (PCA) and XGBoost Algorithms. The proposed model outperforms and produces 97.76% accuracy, 94.51% precision; recall rate is 93.44% and 93.97% F1-Score. This hybrid approaches is better to handle the categorical data and able to find the pattern well. The model outperforms the proposed work in terms of Decision Tree, Gaussian Naïve Bayes, KNN, and Decision Tree, with an improvement of 3.92% and 0.59% improvement in precision, 8.60% and 7.24% improvement in recall, 16.33% and 13.77% in F1-Score, respectively. The integration of ML into IDS marks a significant step towards more intelligent and responsive cyber security frameworks, crucial for defending against the dynamic and increasingly sophisticated threat landscape of today's digital world.

Machine learning-based Intrusion Detection Systems (IDS) have shown promise in improving network security. However, further research is needed to address challenges and enhance their effectiveness. Key suggestions include exploring advanced anomaly detection techniques, developing automated feature engineering techniques, handling imbalanced datasets, real-time processing and scalability, robustness against evasion tactics, explainable AI, integration with other security systems, standardized evaluation metrics, privacy-preserving techniques, adaptive and self-learning systems, and user behavior analysis. Advanced algorithms like deep learning, reinforcement learning, and hybrid models can detect zero-day attacks and novel threats.

Compliance with Ethical Standards:

Funding: No funding was received for conducting this study.

Data Availability: The data and material of the manuscript is available.

Code availability: The code is available in GitHub.

Conflicts of interest: There is no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Louati, F., Ktata, F.B. et al. "Big-IDS: a decentralized multi agent reinforcement learning approach for distributed intrusion detection in big data networks". – In: Cluster Computing, March 2024, Volume 27, pages 6823–6841. https://doi.org/10.1007/s10586-024-04306-9.

[2] Gupta, C., Kumar, A. & Jain, N.K. An Enhanced Hybrid Intrusion Detection Based on Crow Search Analysis Optimizations and Artificial Neural Network. Wireless Pers. Commun. 134, 43–68 (2024). https://doi.org/10.1007/s11277-024-10880-3.

[3] Gupta, N., Jindal, V. et al. "A Survey on Intrusion Detection and Prevention Systems". – In: SN Computer Science. SCI. June 2023, Volume 4, article number 439. https://doi.org/10.1007/s42979-023-01926-7.

[4] Gupta, C., Kumar, A. & Jain, N.K. Intrusion defense: Leveraging ant colony optimization for enhanced multi-optimization in network security. Peer-to-Peer Netw. Appl. 18, 98 (2025). https://doi.org/10.1007/s12083-025-01911-2.

[5] AL-Syouf, R., Bani-Hani, R. & AL-Jarrah, O.Y. "Machine learning approaches to intrusion detection in unmanned aerial vehicles (UAVs). – In: Neural Computing & Application", August 2024 Volume 36, pages 18009–18041. https://doi.org/10.1007/s00521-024-10306-y.

[6] Kumar, V., Kumar, V., Singh, N. et al. "P3IDF-EC: PCA-Based Privacy-Preserving Intrusion Detection Framework for Edge Computing". – In: SN COMPUT. SCI. August 2024. Volume 5. https://doi.org/10.1007/s42979-024-03152-1.

[7] Behiry, M.H., Aly, M. "Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods". – In: J Big Data, January 2024, volume 11. https://doi.org/10.1186/s40537-023-00870-w.

[8] Altamimi, S., Abu Al-Haija, Q. "Maximizing intrusion detection efficiency for IoT networks using extreme learning machine". – In: Discover Internet Things, July 2024, volume 4. https://doi.org/10.1007/s43926-024-00060-x.

[9] Gupta, C., Kumar, A. & Jain, N.K. Intelligent intrusion detection system based on crowd search optimization for attack classification in network security. EURASIP J. on Info. Security 2025, 22 (2025). https://doi.org/10.1186/s13635-025-00205-7.

[10] Ajmal, S., Ashfaq, R.A.R., Raza, A. et al. "IDS-FRNN: an intrusion detection system with optimized fuzziness-based sample selection technique". – In: Neural Computing & Applications. September 2024. https://doi.org/10.1007/s00521-024-10333-9.

[11] Patthi, S., Singh, S. et al. "2-layer classification model with correlated common feature selection for intrusion detection system in networks". – In: Multimedia Tools and Applications January 2024 Volume 83, pages 61213–61238. https://doi.org/10.1007/s11042-023-17781-w.

[12] Al-Haija Qasem A, Saleh E et al. "Detecting port scan attacks using logistic regression". – In: 4th International symposium on advanced electrical and communication technologies (ISAECT), pages 1–5. IEEE. https://doi.org/10.1109/ISAECT53699.2021.9668562.

[13] Zaben, S.O. "IDC-insight: boosting intrusion detection accuracy in IoT networks with Naïve Bayes and multiple classifiers". – In: International Journal of Information Technology June 2024. https://doi.org/10.1007/s41870-024-02026-2.

[14] Al-Haija Qasem A, McCurry Charles D, et al. "Intelligent self-reliant cyber-attacks detection and classification system for IOT communication using deep convolutional neural network". – In: 12th international networking conference: INC 2020 12, pages 100–116. Springer.

[15] Saurabh, K., Sharma, V., Singh, U. et al. "HMS-IDS: Threat Intelligence Integration for Zero-Day Exploits and Advanced Persistent Threats in IoT". – In: Arabian Journal for Science and Engineering, July 2024. https://doi.org/10.1007/s13369-024-08935-5.

[16] Gupta, C., Kumar, A., Jain, N.K. (2023). A Detailed Analysis on Intrusion Detection Systems, Datasets, and Challenges. "Advances in Data Science and Computing Technologies". Lecture Notes in Electrical Engineering, vol 1056. Springer, Singapore. https://doi.org/10.1007/978-981-99-3656-4_26.

[17] Roshan, K. et al. Ensemble adaptive online machine learning in data stream: a case study in cyber intrusion detection system. – In: International Journal of Information Technology, February 2024. https://doi.org/10.1007/s41870-024-01727-y.

[18] Najafli, S., Toroghi Haghighat, A. et al. "A novel reinforcement learning-based hybrid intrusion detection system on fog-to-cloud computing". – In: The Journal of Supercomputing, August 2024, Volume 80, pages 26088–26110. https://doi.org/10.1007/s11227-024-06417-x.

[19] Wang, K., Li, J. & Wu, W. "A novel transfer extreme learning machine from multiple sources for intrusion detection". – In: Peer-to-Peer Networking and Applications. October 2024, Volume 17, pages 33–47. https://doi.org/10.1007/s12083-023-01569-8.

[20] Ngo, VD. Vuong, TC, Van Luong, T. et al. "Machine learning-based intrusion detection feature selection versus feature extraction". – In: Cluster Computing, July 2024, Volume 27, pages 2365–2379. https://doi.org/10.1007/s10586-023-04089-5.

[21] Mustafa, Z., Amin, R., Aldabbas, H. et al. "Intrusion detection systems for software-defined networks: a comprehensive study on machine learning-based techniques". – In: Cluster Computing, April 2024Volume 27, pages 9635–9661. https://doi.org/10.1007/s10586-024-04430-6.

[22] Madhuri, S., Lakshmi, S.V. "A machine learning-based normalized fuzzy subset linked model in networks for intrusion detection". – In: Soft Computing. May 2023. https://doi.org/10.1007/s00500-023-08160-6.

[23] Dubey, S., Gupta, C. (2024). An Effective Model for Binary and Multi-classification Based on RFE and XGBoost Methods. "Intrusion Detection System. Cyber Security and Digital Forensics". Lecture Notes in Networks and Systems, vol. 896. Springer. https://doi.org/10.1007/978-981-99-9811-1_3.

[24] Liu, Y., Zhang, K. & Wang, Z. "Intrusion detection of manifold regularized broad learning system based on LU decomposition". – In: The Journal of Supercomputing, June 2023 Volume 79, pages 20600–20648. https://doi.org/10.1007/s11227-023-05403-z.

[25] Gupta, C., Kumar, A., Jain, N.K. (2025). Optimization Accuracy of Intrusion Detection System Based on Multilayered Neural Network. "Business Intelligence, Computational Mathematics, and Data Analytics. IBCD". Communications in Computer and Information Science, vol 2413. Springer, Cham. https://doi.org/10.1007/978-3-031-87511-3_14.

[26] Wang, X., Dai, L. & Yang, G. "A network intrusion detection system based on deep learning in the IoT". – In: The Journal of Supercomputing July 2024, Volume 80, pages 24520–24558. https://doi.org/10.1007/s11227-024-06345-w.

[27] Merzouk, M.A., Neal, C., Delas, J. et al. "Adversarial robustness of deep reinforcement learning-based intrusion

detection". – In: International Journal of Information Security August 2024 Volume 23, pages 3625–3651. https://doi.org/10.1007/s10207-024-00903-2.

[28] Maseno, Jain, T., Gupta, C. (2022). Multi-Agent Intrusion Detection System Using Sparse PSO K-Mean Clustering and Deep Learning. "International Conference on Artificial Intelligence: Advances and Applications. Algorithms for Intelligent Systems". Springer, Singapore. https://doi.org/10.1007/978-981-16-6332-1_10.

[29] Bhattacharya, S., S, S. R. K., Maddikunta, P. K. R., Kaluri, R., Singh, S., Gadekallu, T. R., Alazab, M., & Tariq, U. (2020). A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU. Electronics, 9(2), 219. https://doi.org/10.3390/electronics9020219.

[30] Amaouche, S., AzidineGuezzaz, Benkirane, S. et al. IDS-XGbFS: a smart intrusion detection system using XGboostwith recent feature selection for VANET safety. Cluster Comput 27, 3521–3535 (2024). https://doi.org/10.1007/s10586-023-04157-w.

[31] Amaouche, S., AzidineGuezzaz, Benkirane, S. et al. IDS-XGbFS: a smart intrusion detection system using XGboostwith recent feature selection for VANET safety. Cluster Comput 27, 3521–3535 (2024). https://doi.org/10.1007/s10586-023-04157-w.

[32] Pourahmad, Zahra, Hooshmand, R.,Madani,S. Mohammad. (2024). "Strengthening of Power Grid Protection Systems Against Cyber-Attacks: A Comprehensive Review" Iranian Journal of Electrical and Computer Engineering.

[33] Abolfazl Sajadi,Bijan Alizadeh, (2024). "SQ-PUF: A Resistant PUF-Based Authentication Protocol against Machine-Learning Attack" Iranian Journal of Electrical and Computer Engineering.

[34] Boshra Pishgoo, Ahmad akbari azirani. (2022). "Improving IoT Botnet Anomaly Detection Based on Dynamic Feature Selection and Hybrid Processing", Iranian Journal of Electrical and Computer Engineering, B- Computer Engineering, Issue 2.

Journal of Information Systems and Telecommunication
Vol.13, No.4, October-December 2025, 300-315
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Compilation of Avatar Development Roadmap in Iranian Banking with the Life Cycle Approach of System Development and Human-Computer Interaction

Amir Bahador Morovat [1*], Farhad Nazari Zadeh [2], Ahmad Haghiri Dehbarez [1]

[1].Department of Industrial and System Studies, Eyvanekey University, Semnan, Iran
[2].Department of Industrial Engineering, Malek Ashtar University of Technology, Tehran, Iran

Received: 03 Mar 2025/ Revised: 04 Sep 2025/ Accepted: 02 Nov 2025

Abstract

The spread and use of emerging technologies have led to a significant transformation in the banking industry and has created widespread changes in the relationship between customers and banks. These changes have led to avatars, previously seen in Hollywood movies, entering the banking sector and are now used as useful tools for providing services to bank customers. Considering the move of Iranian banks towards the adoption of emerging technologies and the willingness of these banks' customers to use these technologies, this research outlines the roadmap of avatar technology in six stages of requirements gathering and analysis, system development, system implementation and coding, testing, deployment and system operation and maintenance, utilizing the expertise of 11 researchers from private and public banks as well as IT and information technology specialists in Iran. In addition, at each stage of the roadmap, the focus has been on customer satisfaction and improving the quality of avatars through human-computer interaction approaches. For this purpose, an estimated timeline of 37 to 49 weeks has been proposed for the roadmap, which describes the necessary actions for each stage, along with possible challenges and issues. What is certain is that the implementation and use of avatars in the Iranian banking industry requires short-term, medium-term, and long-term strategic planning to enable the use of this technology, according to the proposed strategic roadmap.

Keywords: Avatar; AI; Machin Learning; Roadmap; System Development Life Cycle (SDLC); Human-Computer Interaction (HCI).

1- Introduction

The rapid diffusion of digital financial technologies has reshaped value creation and service delivery in banking, pushing institutions toward conversational, always-on interfaces. Among these technologies, avatar-based agents have emerged as a credible vehicle for human-like interaction, personalization at scale, and operational efficiency in front-office and self-service channels [1]. Early adopters report gains in customer engagement and process throughput when avatars are embedded in well-designed service journeys [2, 3].

Despite global advances, avatar deployments in the Iranian banking sector remain limited and largely experimental. Implementations are often interface-centric rather than life-cycle-centric, with insufficient alignment to regulatory, cultural, and infrastructural particularities of domestic banking [4]. This situation elevates the need for a structured, bank-ready roadmap that goes beyond UI novelty to govern requirements, risks, and organizational integration [5, 6].

Prior studies discuss avatars across marketing, virtual environments, and interface design; however, they seldom integrate system development methodologies with user-experience principles for banking contexts. The literature lacks a comprehensive, stage-based roadmap that connects requirements engineering, design, implementation, evaluation, and governance of avatar systems in banks—particularly under Iranian constraints (e.g., data governance, language, service norms) [7, 8]. This gap motivates the present research.

This paper contributes a technology roadmap for avatar implementation in Iranian banking that integrates the System Development Life Cycle (SDLC) with Human–

✉ **Amir Bahador Morovat**
Amir.bahador.19197@gmail.com

Computer Interaction (HCI). Unlike works centered primarily on interface aesthetics or isolated pilots, we articulate bank-grade stages, deliverables, and decision gates that couple user-centered design with enterprise concerns (risk, compliance, and scalability). The roadmap offers actionable guidance for executives and designers while extending the academic discourse on avatarized financial services [9, 10].

Accordingly, we address: How can a comprehensive, user-centered roadmap for avatar technology be designed for the Iranian banking industry to guide end-to-end development and deployment? We aim to specify stages, artifacts, and gates that connect business needs, user requirements, and technical implementation, ensuring regulatory compliance and measurable service outcomes [11, 12].

Our methodological stance derives explicitly from SDLC—to structure complex technology programs—and HCI—to ensure usability, accessibility, and adoption. SDLC supplies the staged governance (from requirements to maintenance), whereas HCI contributes evidence-based principles for interaction design and evaluation[13]. Their integration aligns technical feasibility with human factors throughout the project life cycle.

To enrich analytical rigor in later sections, we propose the following testable propositions that align with our roadmap:

- P1/H1: Adherence to SDLC gates combined with HCI evaluation will be positively associated with customer satisfaction with avatar interactions.
- P2/H2: Organizations that embed user-research insights into requirements and prototyping will achieve higher task completion and containment in avatar channels
- P3/H3: Roadmap-guided deployments will show improved operational efficiency (e.g., shorter handling times) compared with ad-hoc deployments [2, 9].

In addition to addressing an existing gap in the Iranian banking literature, this paper uniquely contributes by presenting an integrated, stage-based roadmap that systematically combines system development methodologies with user experience principles for the first time in this context. The proposed framework offers actionable, locally relevant guidance for banking executives and technology implementers while advancing the academic discourse on technology-enabled service design.

In the subsequent sections of this paper, the research literature and background of avatar technology in financial services are discussed first. Next, a detailed framework is presented integrating the System Development Life Cycle (SDLC) and principles of Human-Computer Interaction (HCI) for the systematic development of avatars in Iranian banking. Each stage of the roadmap is then elaborated with practical actions, challenges, and user experience requirements. Following this, the results and analytical findings are discussed to highlight the practical and comparative aspects of the proposed approach. The paper concludes with a summary of innovations and research achievements, together with a dedicated section offering detailed recommendations for future research directions.

2- Literature Review and Research Background:

2-1- Research Literature:

2-1-1- Avatars: Concepts, Types, and Uses

The term avatar has been framed across multiple traditions—from visual stand-ins and embodied agents to adaptive, context-aware service representatives. [14, 15]. Conceptualizations typically span:

1. visual/representational definitions;
2. contextual/locational accounts across media and cyberspaces;
3. historical/linguistic roots;
4. marketing and identity framings as replicas or personae;
5. temporal/interactivity considerations distinguishing synchronous vs. asynchronous mediation [16-18].

Avatars are now documented across industries—gaming, tourism, education, and public services—gradually converging on service co-production with humans in digital channels [19, 20].

2-1-2- Avatars in Financial Services

In banking, avatars serve as frontline conversational agents, onboarding assistants, and advice companions, mediating tasks such as KYC prompts, eligibility checks, and product guidance. Studies note measurable improvements in engagement, perceived social presence, and guidance quality when avatar behaviors and scripts reflect user intent and financial literacy levels [21,22]. Technology-roadmapping work in adjacent sectors likewise emphasizes staged capability maturation and governance for safety-critical or regulated environments [23-25].

2-1-3- Iranian Context and Early Implementations

Domestic deployments in Iran—often inspired by media mascots and public-sector information campaigns—illustrate cultural receptivity to character-based communication, yet bank-specific avatar programs are scarce and typically UX-led rather than life-cycle-managed [2, 12].

Sectoral surveys highlight infrastructure constraints, language/voice design, and regulatory alignment as pivotal to scale-up [6, 26]. Broader work on persuasive and identity-laden digital characters also signals effects on trust, attention, and purchase intention—implications that banking avatars must responsibly harness [27, 28].

2-1-4- System Development Life Cycle (SDLC)

The SDLC provides a disciplined, stage-based pathway—requirements, design, implementation, testing, deployment, maintenance—supported by decision gates and artifacts that mitigate scope creep and integration risk. Variants (e.g., Waterfall, Iterative, Agile) can be adapted to compliance-heavy banking environments where auditability and change control are essential [29,30]. Prior roadmapping research underscores the value of capability staging and traceability of decisions across the program timeline [46-48]. In our work, SDLC anchors the technology governance of avatar projects [31, 32].

2-1-5- Human–Computer Interaction (HCI)

HCI contributes principles and methods—user research, usability heuristics, accessibility, prototyping, and empirical evaluation—that ensure avatar systems are intuitive, inclusive, and trustworthy. Evidence links HCI-informed design to higher task success, perceived usefulness, and adoption in information systems [48, 49]. Classic and contemporary HCI sources provide foundations for evaluation protocols later employed in our roadmap (e.g., scenario-based testing, think-aloud, SUS-like metrics) [53].

2-1-6- Integrating SDLC and HCI for Avatar Programs

A gap in the literature is the systematic integration of SDLC governance with HCI evaluation for avatar deployments in banking. Existing studies tend to optimize one dimension at a time (e.g., interaction design) without codifying life-cycle artifacts (requirements baselines, UX evidence repositories, acceptance criteria) that carry through to deployment and maintenance [36-38]. We address this by mapping HCI activities onto SDLC stages—ensuring every gate is informed by user evidence and that design intents persist into production and evolution [39, 40].

2-1-7- Positioning vis-à-vis Prior Studies

Compared with prior works that report pilot-level experiences or interface prototypes [21,41], our contribution is a bank-grade roadmap with explicit stage definitions, artifacts, and evaluation hooks that bind user-centric evidence to enterprise risk and compliance controls. This synthesis clarifies where earlier contributions inform our approach (e.g., social presence effects, script design) and where we extend the state of practice (e.g., gate criteria, organizational readiness). This framing prepares the ground for a subsequent discussion section where we will compare our results to similar studies in detail.

2-1-8- Implications for Methodology

The review justifies two design choices in our method: (i) adopting SDLC to manage complexity, traceability, and compliance; and (ii) embedding HCI throughout to safeguard usability and adoption. These choices shape our research propositions (H1–H3) and the evaluation checkpoints later used to qualify the roadmap's effectiveness in banking settings [31, 26, 42].

2-2- Research Background:

Previous studies have shown that the use of avatars improves user experience; however, most existing studies have focused on designing avatars for entertainment or marketing purposes in general, and few studies have examined the use of avatars in the banking field.

Ahmadzadeh, Tabataba'ian, and Shahrestani conducted a study in 2022 to investigate the effect of avatar characteristics on customer identification and purchase intention, with the mediating role of customer involvement in the metaverse. The statistical population of the study was the millennial generation and those born after that in the city of Isfahan. Due to the unlimited statistical population, 384 people were considered according to the Greggsy and Morgan table. The results of the study showed that the mental ability, social skills, athletic ability, artistic/musical ability, and physical attractiveness of the avatar have a positive and significant effect on customer identification. Also, customer identification has a positive and significant effect on customer involvement, customer involvement on purchase intention, and customer identification on purchase intention [28].

In a study based on a theoretical framework of avatar anthropomorphic realism, nonverbal social cues, the eye-mind hypothesis, and interaction process analysis, Lee et al. (2025) examined the effect of avatar gaze behaviors on users' attention allocation and perception. They showed that both avatar gaze type and interaction type significantly affected participants' attention allocation. Natural gaze behavior and task interactions reduced the general pattern of gaze avoidance observed in previous studies. This research emphasizes the importance of the gaze and type of interaction of avatars [39].

In a study conducted by Torabi, Hassangholipour, Yasouri, and Jafari Zare in 2021, the conceptualization and presentation of the avatar marketing model in Iran was discussed. This qualitative research, which is based on the data-driven technique and the Strauss and Corbin approach, was conducted through library studies and semi-structured interviews. The statistical population in this study was experts who, as full professors or associate professors, had many years of teaching experience and were familiar with the new concepts of marketing management and art. The sampling method was also purposeful snowball sampling. The findings of this study identify the main characteristics of avatars in marketing, which are based on two factors:

behavioral realism and visual realism for consumer understanding. They then outline the avatar marketing model with 368 concepts and 38 categories based on six dimensions: causal conditions, context, intervention, main phenomenon, strategies, and consequences. The results of this study indicate that avatars can increase accuracy and precision in customer contact by simulating human behavior in marketing using data science and artificial intelligence, and provide positive and competent performance in unplanned situations and processes to promote the marketing of products and services [41].

Lam (2025) examined the ethical implications of using avatars and virtual reality (VR) in education, focusing on issues such as privacy, identity representation, psychological impact, equity of access, and cyberbullying. This paper proposes stronger and more comprehensive ethical guidelines by integrating Confucian ethics with contemporary ethical frameworks, including epistemology, utilitarianism, and virtue ethics. The holistic approach of Confucian ethics ensures respect for students' identities, mental well-being, and equitable learning opportunities. Ultimately, fostering a culture of virtue, respect, and inclusion can lead to a more ethical and harmonious educational landscape through the responsible use of educational technology. Finally, legal and regulatory approaches to overcome existing challenges are reviewed and examined [40].

In 2024, Bing Hu presented an avatar-based framework for integrated customer identification in banking systems using artificial intelligence and generative graph neural networks. Hu believes that models enhance existing data sets and ensure the completeness and accuracy of customer profiles, so in his proposed model, neural networks are used to simultaneously model complex relationships and interactions in customer data, capture complex dependencies, and increase the accuracy of customer identification. The researcher states that the proposed framework offers significant benefits, including improved regulatory compliance, increased operational efficiency, and superior customer experience. The research also notes that by providing a unified view of customer data, financial institutions can better detect and prevent fraudulent activities, meet stringent regulatory requirements, and provide personalized services. Government regulatory departments can more effectively manage public funds and ensure transparency. Third-party service providers can use customer profiles to better provide services and manage risk [12].

In 2024, Silva and Campos, relying on the TCCM framework, conducted a systematic literature review of research conducted on avatars in reputable journals between 2005 and 2022. This study examines the historical development of the avatar topic, both theoretically and methodologically, examines the key factors involved in avatar marketing strategy, and considers the reasons for studying and using avatar marketing. Finally, an attempt has been made to develop an integrated conceptual framework, a conceptual range of avatar marketing concepts, and to help provide a unified concept for avatar marketing [43].

Miao et al, in 2022 addressed the emerging theory of avatar marketing in a study. In this study, the conceptual and key elements of the term avatar are identified and critically evaluated, and a definition and typology of avatar design elements are presented. The alignment of avatar form realism and behavioral realism among different possible cases, and the effectiveness of avatars, are also examined. Finally, the researchers present an emerging theory of avatar marketing by triangulating insights from the essential elements of avatars, a combination of existing research and business practices. This proposed framework considers key theoretical insights, research propositions, and important managerial concepts for this expanding field of integrated marketing strategy and outlines a research agenda to test the propositions and insights, as well as to advance future research [14].

Although research on avatars has expanded, especially since 2005, attention to this issue has been focused more on topics such as computer games and marketing-related topics. In domestic and foreign research, the concept of avatars in the banking industry and the provision of a roadmap for the use of this technology in the banking industry have not received much attention. Additionally, previous studies have mainly focused on the technical and design aspects of avatars, while human-computer interaction (HCI) and system development life cycle (SDLC) requirements have been less studied. This study fills this gap by providing a comprehensive roadmap and framework for the development and implementation of banking avatars in Iran. Unlike previous studies that have focused more on the marketing and digital interaction aspects of avatars, this study is the first to provide a comprehensive roadmap for the development and implementation of banking avatars in Iran. The combination of the two approaches, SDLC and HCI, in this context has provided a scientific and practical framework for the deployment of this technology in the Iranian banking industry.

3- Research Method

The present research is a descriptive study based on epistemology and future planning of the technology space, and it was conducted based on the opinions of experts and their consensus. For this purpose, a panel of 5 experts active in the field of financial industry technologies was initially formed, and according to the experts' opinions, the system development life cycle approach and the human-computer interaction approach were selected as the methods for drawing the technology roadmap.

To create the technology roadmap, unstructured (in-depth) interviews were conducted with 11 experts in the field of new financial technologies who had at least a master's

degree in fields related to new financial technologies and more than 15 years of experience in banks or financial service companies. These experts were purposefully selected. Prior to conducting the interviews, the results of library studies on avatars, the system development life cycle approach, and the human-computer interaction approach were shared with the interviewees. Individual interviews were then conducted with each expert.

Before starting the interviews, the scope and context of the problem were explained to the experts by the researcher. Four sessions were held with each expert at approximately one-month intervals, with questions raised about each stage separately. Each interview lasted 45 to 70 minutes. The number of interviews was based on a saturation pattern, and the content analysis method with open coding was used to analyze the interviews. At each stage, the concepts from library studies were also used to complete the stage. At the end, a roadmap was compiled for validation and approval by each of the 11 research experts, and their corrective points were considered.

According to the Expert Panel, the system development life cycle approach was considered for drawing the technology roadmap, and the waterfall model was chosen among the various existing models. The waterfall model is simple, methodical, and easy to understand and implement. It works well and produces correct results. The main advantage of the waterfall model for the system development life cycle is that it provides a structure for organizing and controlling a system development project. However, the most important methodological requirement in this approach is the accurate identification of user needs [26]. Due to its regular and linear structure, this model is suitable for large and complex projects such as avatar development. In general, the waterfall model of the system development life cycle has six stages as follows:

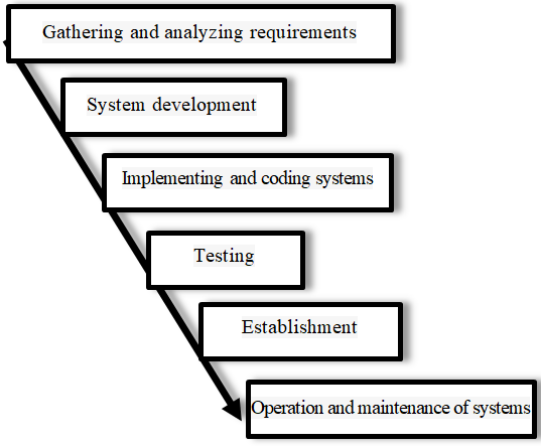

Fig. 1 Stages of the waterfall model in the system development life cycle [49]

Another approach in this research is human-computer interaction, which deals with how humans and computers communicate and examines the principles of designing, evaluating, and implementing interactive systems with the aim of creating a user-friendly and effective experience. There are various models for describing and analyzing human-computer interaction, each focusing on specific aspects of this interaction. The user experience (UX) model was used to draw a technology roadmap in the Iranian banking industry. This model, focusing on the user, improving customer satisfaction, and achieving the goal [42], is considered a suitable option for drawing an avatar technology roadmap. In addition, this model is compatible with the waterfall method of the system development life cycle. The user experience model focuses on strategy, scope, usability structure, and visual values. While describing each of the components, the research experts were asked to pay attention to each of the six stages of developing a technology roadmap in applying these items. Fig 2 presents a diagram of the research process.

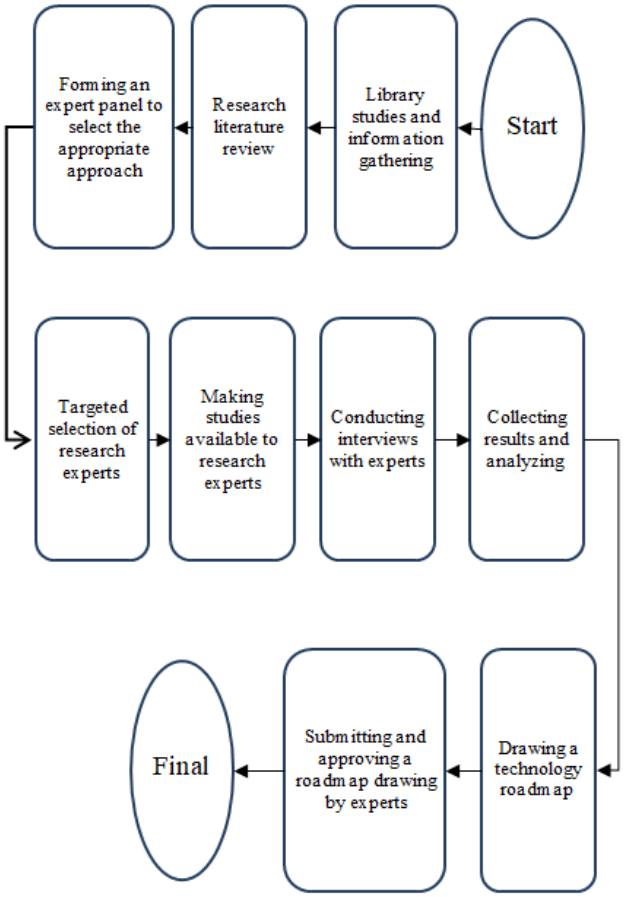

Fig. 2 Research implementation process

4- Implementation Steps:

In this research, the avatar technology roadmap has been drawn according to the six-step approach of the waterfall model of the system development life cycle. According to the opinions of research experts and library studies, the actions required at each stage and the existing challenges, along with the proposed duration, have been considered. Additionally, the human-computer interaction user experience approach has been addressed at each stage.

4-1- Requirement's Collection and Analysis:

According to the research experts, this stage is considered one of the most important stages of drawing the avatar technology roadmap in the banking industry because the results obtained from it will be used in the later stages of the project. Additionally, with a detailed analysis, it is possible to identify customer needs and the technical limitations of the bank and take action in designing an effective and compatible avatar.

Avatars can be used as powerful tools to solve many common customer problems and improve user experience. One of the basic needs of bank customers is addressing questions and ambiguities regarding receiving services or carrying out financial transactions (questions about the conditions of the facilities, how to transfer, etc.), which typically involve visiting bank branches and standing in queues or calling support centers and spending time waiting for expert responses. While meeting this need will be easily resolved through avatars, avatars can quickly understand complex questions and provide accurate answers 24/7 from different portals (branch, mobile banking, internet banking) using natural language processing algorithms. They can act as smart assistants, helping customers quickly find the information they need.

Avatars can help customers perform simple transactions such as transferring funds, paying bills, etc., and act as financial advisors to help customer's select appropriate products and services. Additionally, by using a simple and smooth user interface, avatars can make the transaction process easier for customers. In cases where access to bank branches is not possible, avatars can act as virtual branches, allowing customers to access banking services at any time and place, thereby increasing customer satisfaction and productivity for the bank.

Other important considerations at this stage include analyzing and identifying customer behavior patterns and paying attention to different groups of bank customers. Factors such as the desire to visit bank branches in person, uncertainty about the security of technologies, lack of awareness among some bank customers, and, in some cases, the costs of using technologies like avatars are crucial in determining customer behavior patterns. By identifying these patterns, it is possible to define various user personas, accurately identify the needs of each customer group, and design a simple, intuitive, and understandable avatar for all customer groups and every age group. Additionally, identifying customer behavior patterns in designing conversations is effective in gaining customer trust and providing personalized services to customers.

For example, a young customer may seek fast, efficient, and online banking services and be interested in new technologies. A middle-aged customer may seek financial advice and require accurate and reliable information. An elderly customer may seek simple and understandable services and need more help using an avatar. Addressing these issues will increase the acceptance of avatar technology by bank customers, improve user experience, reduce costs by automating many processes, and increase revenue by providing better and personalized services.

By examining the above issues, it is possible to identify the capabilities required by the avatar, such as answering frequently asked questions, guiding customers in performing simple transactions, or authenticating identities. These capabilities can be designed based on a needs analysis and a review of customer behavior patterns.

Another important issue at this stage is identifying the existing infrastructure necessary for avatar technology in the bank. The use of avatar technology in banking requires various technologies, such as natural language processing, speech recognition, machine learning, artificial intelligence, virtual and augmented reality, user interfaces, and cybersecurity. The existence or use of these technologies should be considered for the implementation of avatar technology. Additionally, the availability of a suitable infrastructure for different communication channels to provide avatar services, the data and information infrastructure available in the bank, and the possibility of integrating avatar communication with the existing infrastructure in the bank are factors that should be considered. Addressing challenges such as customer acceptance of technology, along with the appropriate and required infrastructure, can contribute to the successful implementation of the project.

Considering the human-computer interaction approach, user research and understanding are important. Identifying the needs, goals, and different behaviors of users through tools such as interviews, questionnaires, and observations can help us gain better insights. Additionally, drawing a user journey map at this stage can be valuable, as it outlines the user's interaction path with the avatar from beginning to end. This map helps us identify user touchpoints with the avatar and pinpoint weaknesses and opportunities for improvement.

Another essential step in this phase is to identify and document the security and privacy requirements for the Avatar system. This includes identifying the type of sensitive data, potential threats, and relevant laws and regulations. Potential legal barriers to the development and operation of the Avatar system should also be identified and reviewed, which include laws related to the provision of

banking services, data protection, and other relevant regulations. At this stage, the needs and expectations of users should be identified and analyzed to ensure their acceptance of the Avatar system. It is important to examine the factors affecting the acceptance of technology by users and design the system in a way that meets their needs.

Therefore, according to the experts' opinions, a SWOT analysis can be conducted to draw a technology roadmap in the Iranian banking industry with the aim of identifying internal and external factors affecting the implementation of avatar technology. SWOT is a powerful tool for identifying internal strengths and weaknesses and external opportunities and threats.

Investigating and identifying internal and external opportunities and threats will greatly contribute to analyzing the gap of avatar technology in the Iranian banking industry. Based on interviews with research experts and library studies, the gap in this technology in the Iranian banking industry was analyzed. The purpose of this analysis is to identify the gaps between the current state of Iranian banking and the desired state (full use of avatar technology). Accordingly, the inadequacy of the technical infrastructure in Iranian banking for implementing avatar technology, the absence of specific laws and regulations for the use of avatar technology in Iranian banking, the lack of familiarity and trust of many customers with avatar technology, and the lack of expertise required to develop and implement avatar technology in Iranian banking can be considered the main gaps in the use of this technology in the Iranian banking industry.

Considering these gaps, investing in the development of technical infrastructure, formulating laws and regulations, holding educational and culture-building programs to familiarize customers with avatar technology, and attracting experts in the field of avatar technology and training existing personnel can play an effective role in filling the existing gaps.

Additionally, according to the research experts, the level of advancement of avatar-related technologies (such as artificial intelligence, natural language processing, etc.), the level of use of avatar technology in other industries and countries worldwide, the existence of the necessary technical infrastructure for implementing avatar technology in Iran, the existence of specific laws and regulations for the use of avatar technology in Iran, and the level of familiarity and acceptance of avatar technology by customers in Iran are considered criteria for the maturity of avatars in the Iranian banking industry. Accordingly, the maturity of avatar technology in the Iranian banking industry is not in a favorable state.

A noteworthy point at this stage is the identification and documentation of all requirements. Considering the above and analyzing the issues raised, decisions can be made regarding the channels used for avatars, the capabilities implemented in them, and how the services provided are delivered. It is advisable to focus on the basic needs of customers and the simple capabilities of avatars at this stage. Over time and with the improvement of infrastructure, more complex capabilities can be added to avatars. According to experts, the suggested duration for this stage is 6 to 8 weeks.

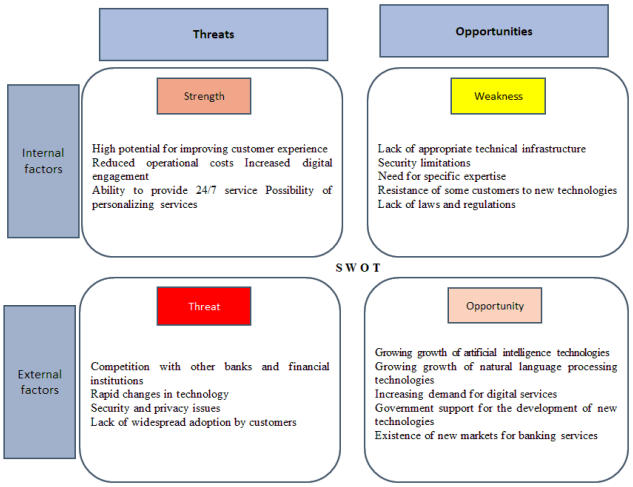

Fig. 3 SWOT Avatar in the Iranian Banking Industry (Source: Researcher's findings)

4-2- Systems Development

This stage will be carried out according to the analysis and identification conducted in the previous stage and involves transforming the initial designs and identified requirements into an operational system, establishing the technical foundations of the avatar. In this stage, the architecture of the Avatar Bank system is designed, which means determining the overall structure of the system, its components, and how the system components interact with each other. In simpler terms, this stage is like designing a building plan, where all details are specified before construction begins. The system architecture determines the overall project path, and by separating the system into smaller components and defining the relationships between them, the complexity of the project is reduced and the possibility of future system development and changes is provided. By specifying the roles in carrying out the project, the members of the development team can effectively collaborate, facilitating cooperation.

The components of the avatar system vary depending on the capabilities of the avatars, but typically include a conversation engine for natural language processing, response generation, conversation management, a database for storing user information, products, services, conversations, a user interface for user interaction with the avatar, channels through which the avatar communicates with users (such as Internet banking, mobile banking, branches), an authentication system, and a bank payment system for performing transactions. In this stage, the

relationships between the components and the method of communication are determined, communication protocols are established, and architectural diagrams are designed to display the relationships between the components.

Selecting appropriate programming languages, frameworks, databases, and development tools, selecting a cloud or local platform to host the system, designing a database structure to store the information required by the system, data normalization to prevent redundancy and improve efficiency, and designing security systems to protect user data and prevent cyber-attacks are other issues that should be addressed in the system architecture.

Based on the project needs, the skills of the development team, and the existing infrastructure, the appropriate programming language and framework are selected. This choice has a direct impact on the performance, maintainability, development speed, and costs of the project. Depending on each portal used (web applications, mobile applications, or back-end systems), different languages and frameworks will be more suitable for use. Languages and frameworks that produce clean and readable code make it easier to maintain and develop the avatar in the future. For example, in web application development, Python with Django is suitable for fast and scalable back-end development, JavaScript with React is suitable for building interactive and dynamic user interfaces, and Node.js is suitable for building server-side and real-time applications. For mobile application development, Swift can be used for developing iOS applications, and Kotlin for developing Android applications. React Native is also suitable for developing iOS and Android applications simultaneously.

According to research experts, Python is suitable for avatar design due to its ability to support the required technologies, such as artificial intelligence, its suitable speed, and high security. Additionally, using this programming language will reduce costs.

Although issues such as infrastructure limitations, changes in identified requirements, and system complexity cause problems in designing the appropriate system architecture, with focus and accuracy in proper design, we can prevent problems in the later stages of development and reduce development costs.

A suitable architecture guarantees system performance, scalability, maintainability, and security. Given the infrastructure limitations in Iranian banks and the need to interact with existing bank systems, the architecture should be designed in such a way that it can easily communicate with core banking systems, customer relationship systems, and other bank systems. The architecture should ensure that the avatar's performance does not suffer if the number of users and data volume increases.

Another important point in system architecture is information security in the avatar. The architecture must protect sensitive user information from security threats.

Information security in the banking system is of great importance. In the implementation phase of the bank avatar, special attention should be paid to security by design, and the necessary security measures should be taken at all stages of system development and implementation. Using strong authentication methods to prevent unauthorized access to user information and encrypting sensitive user data in the database and when transferring information, along with using firewalls, intrusion detection systems, and other security tools to prevent hackers from penetrating the system and performing regular security updates to remove possible vulnerabilities, are some of the things that should be given special attention.

Another important step at this stage is choosing the right platform and natural language processing engine as a crucial step in the development of the bank avatar. In choosing a platform, factors such as the scale of the project, budget, complexity of the desired interactions, and the existing infrastructure in the bank should be considered. A suitable platform should be able to support multiple channels and have the ability to interact with users through different channels, be customizable and able to customize the appearance and behavior of the avatar to suit customer needs, be able to integrate with existing systems, have a high level of security to protect user information, and be scalable given the increasing number of users and the complexity of interactions. The avatar should be simple, intuitive, and attractive so that users can easily interact with it. Choosing the right colors, fonts, and graphic elements is important to create an attractive user interface; the visual design should be such that it displays well on different devices, while ensuring that people with special needs (such as people with disabilities) can also use the avatar.

Natural Language Processing (NLP) Models are essential components of the avatar, responsible for understanding and interpreting the natural language of users. There are various natural language models that can be used in the bank avatar. Natural language processing models include rule-based models, which operate on a set of predefined rules and patterns and are suitable for simple tasks such as recognizing keywords and specific phrases, machine learning models that learn patterns in natural language using training data—these models are suitable for more complex tasks such as emotion recognition, text summarization, and language translation—and deep learning models that use deep neural networks to process natural language. These models are capable of performing very complex tasks such as understanding the meaning of sentences and answering questions. Choosing the right model for a bank avatar depends on various factors such as the complexity of the tasks, the amount of training data available, and the available computing resources.

Another requirement that should be considered at this stage is the database. The database contains information about the bank's products and services, rules and regulations, frequently asked questions and their answers,

and any other information that the avatar needs to respond to users. The database structure should be designed to provide quick and easy access to information. Information in the database can be stored in structured (such as tables) or unstructured (such as free text) forms. To ensure that the information is up-to-date, a system should be created to manage and continuously update the database. Choosing the right database is very important, depending on the type of data, its volume, and how it is used. Experts have chosen PostgreSQL as a suitable option due to its ability to store and retrieve large amounts of data, speed and efficiency, scalability, security, and cost, and have emphasized its low cost and expandability.

The Application Programming Interface (API) is another area that should be considered in this section. It will be used to access information in the bank database such as customer information, account balances, and to perform banking transactions such as money transfers, bill payments. APIs will also be used to collect user feedback and improve avatar performance, develop an admin panel, manage users, manage the knowledge database, update information, monitor avatar performance, and identify potential problems.

It is important to consider challenges such as infrastructure limitations and technical complexities at this stage, while data scarcity and the need to coordinate with existing systems are also important factors.

At this stage, the security and privacy requirements identified in the previous stage should be included in the system design. It is important to pay attention to the use of secure architectures, encryption protocols, and access control mechanisms. Legal solutions should be designed and implemented to overcome the obstacles identified in the previous stage, which could include changes in system design, obtaining the necessary licenses, or changes in the way services are provided. The system should be designed in a way that meets the needs and expectations of users and provides a desirable user experience. This includes designing an appropriate user interface, providing the necessary training, and creating user feedback mechanisms. According to the human-computer interaction approach, designing a simple, intuitive, and attractive user interface that is tailored to the needs and abilities of users should be the criterion for action at this stage. Attention should be paid to designing natural and smooth interactions between the user and the avatar in the design of conversations and animations.

According to research experts, the suggested time for this stage is 8 to 10 weeks. Additionally, attention to customer data security should be a priority, and various tests should be conducted during the development process to ensure the proper functioning of the system. Given the infrastructure limitations of banks, it is better to focus on the basic capabilities of the avatar first and gradually add more complex capabilities over time.

4-3- Implementation and Coding of Systems

According to experts, in the implementation and coding of systems stage, the designs created in the second stage will be transformed into a usable product, with programming codes written and various components of the system connected to each other.

The type of project, the team's skills, and the performance of the programming language are among the factors that influence the choice of these factors. Frameworks provide a predefined structure for the project, saving time and effort, and allowing us to benefit from the experiences of others.

Developing the user interface is another key factor at this stage. The user interface is the user's first interaction with the avatar and plays an important role in the user experience. A good user interface should have an attractive visual design that is consistent with the brand, compatible with a variety of devices and browsers, and usable by users with specific needs. To develop a user interface, it is essential to create initial designs using design tools, build user interface components using the selected language and framework, and connect the user interface components to business logic.

Another factor to consider at this stage is the implementation of the conversation engine based on the selected models. The conversation engine is responsible for understanding and answering user questions. Types of conversation engine models include rule-based models that respond based on a set of predefined rules and patterns, and machine learning models that use machine learning algorithms to learn from training data and produce natural responses. To develop a conversation engine, factors such as collecting a large dataset of questions and answers, preprocessing data to remove noise using natural language processing, selecting the appropriate model based on the type of data and project needs, and training and evaluating the model can be used.

Integration with existing systems and connecting the avatar to various bank systems, such as the central banking system, customer relationship system, etc., through user interfaces should also be considered because it provides access to various information and services.

At this stage, the design and implementation of the database to store user information, conversations, and system settings should also be addressed. This involves identifying the main entities in the system, defining the relationships between entities, and creating a diagram to display the database structure. Security and privacy requirements should also be considered in coding and implementing the system, with the use of secure coding methods, implementation of security mechanisms, and conducting security tests. It is necessary to ensure that the legal solutions designed in the previous stage are implemented in the system.

Coordination between teams and change management are among the challenges of carrying out this stage, while the

complexity of the system may also create problems in implementation.

According to the human-computer interaction approach, the system implementation should be carried out with due care based on the designs made. At this stage, the system should be implemented in a way that users can easily use it and meet their needs.

The suggested time for the third stage, according to the research experts, is 14 to 18 weeks. It is recommended that after the development of each section, testing should be carried out to ensure the correct functioning of that section, and an effective communication system should be established between the design, development, and testing teams.

4-4- Testing

This stage is one of the most sensitive and vital stages in the system development life cycle and ensures that the developed system works correctly, complies with the initial requirements, and can provide a desirable user experience for customers. In the banking industry, which interacts directly with customers, the quality of tests is very important. The start of this stage requires the completion of the previous stages.

Due to the complexity of human-computer interaction and the importance of user experience, it is recommended that testing be performed continuously and throughout the different stages.

This stage should use unit tests to check the correct functioning of each part separately, ensuring that each component processes the inputs and produces the correct outputs. It is also important to check the correct functioning of the various components of the avatar when working together, ensuring that the different components communicate correctly and share data. It should also be tested to ensure the performance and functional evaluation of the avatar under different loads, such as a large number of users simultaneously, and to ensure that the avatar can meet the needs of users in different situations. It should also be tested to assess the ease of use of the avatar by users, ensuring that the user interface is attractive, intuitive, and user-friendly, and to assess the security vulnerabilities of the avatar to ensure that sensitive user information is protected.

Other areas that require testing include evaluating the avatar's behavior under abnormal conditions, such as system errors and network outages, to ensure that the avatar responds correctly to errors. End-user evaluation of the avatar to confirm that their needs are met is also essential.

In this phase, to validate the roadmap and before full deployment, the avatar is piloted on a small scale, and the results are reviewed. This work reveals strengths and weaknesses and can play a key role in the full deployment of the avatar.

Since the user experience approach of human-computer interaction has also been addressed in the technology roadmap in this research, user-centered testing should also be considered. Testing with real users to collect feedback on user experience, using various methods such as interviews, observations, and questionnaires, as well as measuring indicators such as task completion time, error rate, user satisfaction, and avatar usage, and using tools to track user behavior in interacting with the avatar and identify strengths and weaknesses should be on the agenda.

Security and privacy challenges are of particular importance, and the avatar must be tested for security and privacy. These tests include penetration testing, vulnerability testing, and compliance testing with relevant laws and regulations.

The complexity of human-computer interaction, continuous changes in artificial intelligence, and weak technology infrastructure are factors that may disrupt the implementation of this stage. In this stage, the system is tested by real users to ensure their acceptance. A kind of simulation of the technology adoption model is proposed to determine the impact of different policies on behavior change [50].

According to research experts, the general goal of this stage is to identify and fix potential problems before deploying the system in a real environment. Conducting various tests such as unit testing to check the correct functioning of each part of the system (such as the conversation engine, user interface, database, etc.), integration testing to check how different parts of the system interact with each other, system testing to check the overall performance of the system and its compliance with the defined requirements, user-centered testing to evaluate the user experience and user satisfaction with the system, security testing to check system security vulnerabilities and protect user information, and performance testing to check system performance under different loads (such as a large number of simultaneous users) is essential at this stage. Also, in order to use the test results in the next stages, they must be documented. The suggested time for this phase is 6-8 weeks and includes the need to design detailed and comprehensive tests to evaluate all aspects of the interaction, the need to continuously update the tests with new developments in the field of artificial intelligence, the need to design tests that take into account the limitations of the technology infrastructure, and establish a strong relationship between the test team and the development team to ensure high quality.

4-5- Establishment (development)

This phase is the transition of a software system from the development environment to the production environment. In simple terms, this phase is where the avatar is transferred from the laboratory environment to the real environment of the bank and made available to customers. Successful deployment indicates that the project has achieved its goals and is ready to serve customers. If the

system is deployed correctly, it directly affects the initial user experience and reduces the risks of technical problems and service disruptions.

When deploying the system, attention should be paid to issues such as installing software, setting up the database, and configuring the network, implementing security measures to protect customer data and prevent cyber-attacks, providing the infrastructure needed to support the avatar, transferring information from the development environment to the production environment, and ensuring data compatibility with the new environment. This ensures that the avatar works properly in the production environment and meets customer requirements.

Considering the user experience approach of human-computer interaction, attention to such aspects as ensuring that the user experience is seamless and smooth at all stages, including the registration and login process, providing technical support and guidance for new users, and collecting user feedback after deployment for continuous improvement of the system are also essential at this stage.

Training employees on how to use and support the system and informing customers about new capabilities and encouraging them to use the avatar should also be considered.

In deploying the avatar, challenges such as infrastructure complexity, information security, and organizational changes should also be considered. Additionally, the necessary security measures should be taken to deploy the system in an operational environment, which includes secure configuration of servers, installation of firewalls, and other security measures. At this stage, the necessary permissions for operating the system should be obtained from the competent authorities.

The avatar should be deployed in a way that users can easily access and use it. The duration predicted for this stage by research experts is 3 to 5 weeks. Monitoring the performance of the avatar in the early days and resolving any potential issues, preparing a detailed deployment plan (including timing, responsibilities, and resources required), conducting comprehensive pre-deployment tests to ensure proper system performance, and working closely with development, network, security, and support teams are some of the things that should be considered.

4-6- Operation and Maintenance of Systems

At this stage, the system is fully available to users and is used continuously. The main goal at this stage is to ensure the correct and stable operation of the avatar, resolve potential issues, and continuously improve it.

After the avatar is deployed, its performance should be monitored. Continuous monitoring of key performance indicators (KPIs) such as response time, error rate, and capacity is of particular importance. Identifying technical problems and resolving these problems proactively should be considered. Performance indicators such as customer engagement rate, reduction in branch visits, avatar response accuracy, and the impact on customer satisfaction are considered to be the most important key indicators of avatar performance.

At this stage, in addition to training users to use the avatar optimally, support should be provided to users in case of any problems or questions.

Another factor emphasized at this stage is to review the avatar's performance in order to make the necessary changes and modifications to improve it, so that new features can be added if needed by users and to accommodate possible changes in the banking industry.

Given the importance of data security in the banking industry, protecting avatar data against security threats is crucial, and security updates should be made regularly. Periodic data backups can also be useful for recovering data in case of serious problems.

Considering the conditions and characteristics of the Iranian banking industry, according to the research experts, the changing needs of Iranian bank customers and the need to adapt the avatar to these needs, as well as changes in technology due to the rapid growth of technology and the lack of appropriate technology infrastructure in Iranian banks, are among the most important challenges and problems of this stage. These challenges, along with the problems in the complex interactions between users and the avatar, have made this stage difficult. The avatar must be continuously monitored in terms of security and privacy, and corrective measures must be taken if necessary. It must also be continuously monitored for compliance with relevant laws and regulations, and necessary changes must be made as required.

Considering the user experience approach of human-computer interaction, it should be noted that collecting user feedback from using avatars and using this feedback to improve the user interface and performance of the avatar is crucial. Personalized support should also be provided to users based on their specific needs.

4-7- Proposed Technology Roadmap:

The six steps for developing the avatar technology roadmap in the banking industry were determined through the six-stage waterfall model of the system development life cycle, in alignment with expert insights and the user experience principles of human-computer interaction. Figure 3 illustrates the practical implementation timeline for each stage, while the main required actions, associated challenges, and user experience considerations for each step are summarized within the discussion provided in this study.

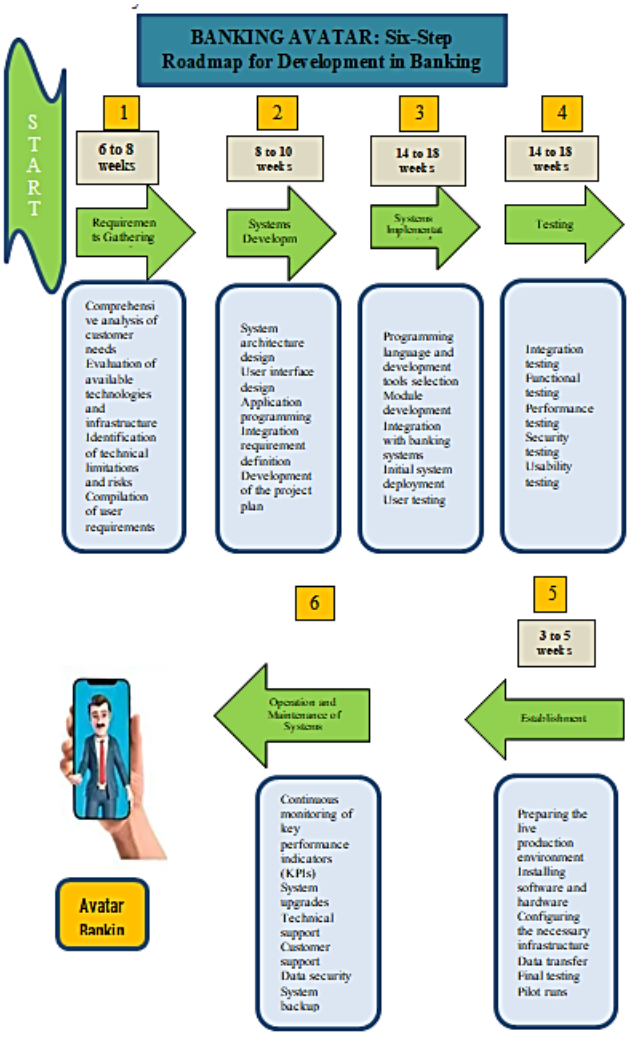

Fig. 4 Proposed roadmap for avatar technology (Source: Researcher's findings)

The implementation and use of avatars in the Iranian banking industry require planning for short-term, medium-term, and long-term strategies to benefit from this technology, according to the proposed roadmap.

Short-term strategies are specific and implementable actions for periods of 3 to 6 months. These include launching the initial version of the avatar and evaluating its performance, implementing the MVP (minimum viable product) version with limited features for testing in a real-world environment, testing user interactions with the avatar, receiving feedback, and optimizing technical and security infrastructures. Key performance indicators (such as customer satisfaction) are needed to measure progress in short-term periods. For this purpose, implementing a pilot version of the avatar technology in a limited range of banking services to evaluate its effectiveness and receive customer feedback should be considered. Developing and implementing a pilot version of the avatar (for example, to guide customers on the website or application), selecting a small group of customers to participate in the pilot test, and collecting quantitative and qualitative data on the customer experience (through surveys, interviews, and recording customer activities) are essential. Finally, analyzing the data and providing a preliminary report on the test results to identify the strengths and weaknesses of the pilot based on customer feedback within a 3- to 6-month period should be carried out.

Medium-term strategies are specific and measurable goals for periods of 6 to 12 months. These strategies are directly derived from long-term goals. Strategies such as improving and developing the banking avatar, increasing the capabilities of the avatar (adding advanced natural language processing, recognizing customer emotions, and intelligent responses), expanding platforms (providing avatar services in mobile applications of banks and banking kiosks), and evaluating the impact of the avatar on the bank's performance (reducing in-person visits, improving the customer experience, increasing productivity) are of interest. Planning medium-term strategies with the aim of improving and developing the pilot version of the avatar based on customer feedback and increasing its capabilities to improve the bank's performance (e.g., reducing costs, increasing customer satisfaction, increasing efficiency). Improving the design and user interface of the avatar based on customer feedback and adding new capabilities to the avatar (e.g., the ability to answer complex questions, perform simple financial transactions) are essential.

Implementing the avatar in other communication channels (e.g., including the avatar in bank branches), along with measuring the impact of the avatar on the bank's performance (e.g., reducing customer waiting time, increasing customer satisfaction, reducing customer service costs), are among the programs that can be implemented in the medium-term strategy. These will be implemented by providing a full report on the results of the avatar improvement and development within a period of 6 to 12 months.

Lon-term strategies are considered for 1 to 3 years. The long-term strategy includes major goals such as integrating and developing advanced capabilities, adding augmented reality and virtual reality technologies for remote banking consultations, integrating with other banking systems and smart trading robots, and complying with international standards to enter advanced digital banking. Other pillars of long-term strategies include vision and scope. According to experts, the vision and scope of using avatars in the Iranian banking industry by 2030 is to use smart avatars to provide a personalized, fast, and user-friendly experience for Iranian bank customers. This technology will lead to a 30% reduction in in-person visits, a 50% increase in customer satisfaction, and a 40% optimization of banks' operating costs.

In accordance with the planning for long-term strategies and achieving the vision of using avatar technology in the Iranian banking industry, action should be taken to add augmented reality technology and enter digital banking to create innovative experiences in digital banking and increase customer satisfaction by 2030. According to the planning, in a 3- to 5-year period, research and development should be carried out in the field of integrating augmented reality technology with avatars, developing and implementing new features using augmented reality technology (e.g., 3D display of customer accounts, guiding customers in bank branches), developing and implementing digital banking solutions using avatars (e.g., performing complex financial transactions with the help of avatars), measuring the impact of avatar technology on customer satisfaction and bank performance by 2030, and presenting a final report on the results of implementing the long-term strategy. Attention should be paid to the vision of drawing customer satisfaction through providing innovative and personalized experiences using avatar technology and augmented reality. Another point in implementing the planned plan is to pay attention to the challenges ahead of the avatar technology roadmap in the banking industry. Obstacles and challenges such as infrastructure and security restrictions due to weak internet and data security in some Iranian banks can be overcome by using secure cloud servers [62] and advanced encryption technologies. The challenge of user acceptance and culture building is also of concern because some customers, especially the elderly, may not trust avatars. For this purpose, educational and digital marketing programs will be useful to familiarize customers with this technology. Legal and regulatory challenges and the lack of specific rules in the field of financial interactions with banking avatars can be overcome by creating banking laws and standards.

5- Discussion

The proposed avatar development roadmap, grounded in SDLC and enriched with HCI principles, outlines a stage-based framework for conceptualizing, designing, deploying, and maintaining avatar services in Iranian banking. Each stage—requirements analysis, design, implementation, evaluation, and maintenance—is linked to specific deliverables and decision gates that combine technical rigor with user-centric evidence [14].

Compared to earlier avatar deployment studies in financial services [44], our results demonstrate greater alignment between interaction design and organizational integration. Prior works have largely emphasized interface appeal and novelty; by contrast, our roadmap formalizes enterprise-level governance, regulatory compliance checkpoints, and integration of user evidence at each gate.

Our findings substantiate the theoretical claim that integrating HCI within SDLC enhances both usability and operational outcomes. The dual-framework approach ensures that human factors are embedded in technical project controls, reducing the gap between conceptual prototypes and production-ready systems.

For practitioners, the roadmap offers actionable guidance on stage sequencing, role allocation, and artifact creation. Banks can adopt this framework to improve avatar adoption rates, streamline change management, and enhance customer satisfaction metrics. By embedding evaluation protocols into each stage, project teams can proactively identify and resolve user-experience issues before full-scale rollout [45].

This study, while comprehensive in its conceptual design, is limited by its reliance on secondary literature and expert validation. Empirical measurement of roadmap-guided deployments in live banking contexts remains a future task. Contextual constraints—such as evolving Iranian banking regulations and digital infrastructure maturity—also delimit the generalizability of our findings.

6- Conclusions

This study established a structured and context-aware roadmap for avatar technology implementation in the Iranian banking sector, integrating proven principles of the System Development Life Cycle (SDLC) and Human-Computer Interaction (HCI). The research process involved comprehensive literature review, expert consultations, and in-depth analysis of local banking needs and constraints. Key development stages—including requirements gathering, design, implementation, testing, deployment, and maintenance—were thoroughly defined, and actionable guidelines for each phase were presented with a focus on user experience and operational feasibility. Collaboration with banking and IT professionals ensured practical alignment and relevance throughout the framework.

The main research innovations and unique contributions of this paper are:

1. Development of an integrated SDLC–HCI framework for avatars: Introducing a systematic approach that aligns the technical progress of banking avatars with rigorous usability standards and user-centric practices, tailored specifically for Iranian banks.
2. Precise definition of each stage and its deliverables: Providing clear operational guidelines, decision gates, and performance measures for every roadmap stage to facilitate effective implementation and project governance.
3. Addressing Iran-specific banking challenges: Adapting all stages of the roadmap to domestic banking regulations, infrastructural realities, and

cultural requirements, ensuring readiness for real-world deployment.
4. Strategic vision for phased development: Articulating distinct short-term, medium-term, and long-term pathways for the scalable and sustainable evolution of avatar-based services in Iranian banks.
5. Direct practical applicability: Offering the banking sector a tangible blueprint to accelerate digital transformation, improve customer engagement, and enhance overall service delivery without reliance on artificial intelligence technologies.

By coupling solid methodological foundations with a robust, locally tailored strategy, this research bridges academic rigor and industry needs—enabling Iranian banks to confidently pursue avatar adoption while addressing the unique operational and cultural challenges of their environment [51,52].

7- 7- Directions for Future Research and Practical Development

Looking ahead, the continued evolution and successful integration of avatar technology in Iranian banking depend on targeted research and strategic development efforts. Although the current study delivers a comprehensive roadmap and practical framework, further exploration is necessary to extend its impact, address real-world complexities, and unlock new opportunities for innovation in banking services. The following areas are recommended for future study and action:

1. Empirical implementation and evaluation: Pilot deployments of the proposed roadmap in selected banks, with systematic measurement of customer experience, operational results, and user acceptance, will validate and refine the framework under real conditions.
2. Enhancement of service integration: Investigating methods to connect avatar solutions with diverse banking systems and customer service platforms, supporting more seamless and efficient operational workflows.
3. User feedback and iterative improvement: Establishing routines for collecting and acting upon end-user feedback to ensure the avatar's ongoing usability, reliability, and alignment with evolving banking needs and customer expectations.
4. Cultural and trust-building initiatives: Conducting targeted studies on the social and cultural factors that influence customer trust and acceptance of avatars, and developing evidence-based strategies for digital literacy and technology adoption.
5. Sector-wide adaptation and expansion: Assessing opportunities to apply the roadmap to related industries—such as insurance or government services—to demonstrate its value beyond the banking sector and foster broader digital transformation.

These future research directions will require coordinated collaboration among technology experts, banking practitioners, and regulatory bodies. Their successful pursuit will further strengthen the foundations for effective, user-centered digital banking services in Iran—without reliance on artificial intelligence technologies.

Appendix

There are no appendices included in this article.

Acknowledgments

The cooperation of the experts used in this research is appreciated.

References

[1] De Brito Silva, M. J., & De Oliveira Campos, P. (2024). Past, present, and future of avatar marketing: A systematic literature review and future research agenda. Computers in Human Behavior: Artificial Humans, 2(1), 100045. https://doi.org/10.1016/j.chbah.2024.100045

[2] Oh, H. J., Kim, J., Chang, J. J., Park, N., & Lee, S. (2023). Social benefits of living in the metaverse: The relationships among social presence, supportive interaction, social self-efficacy, and feelings of loneliness. Computers in Human Behavior, 139, 107498. https://doi.org/10.1016/j.chb.2022.107498

[3] Mystakidis, S. (2022). Metaverse. Encyclopedia, 2(1), 486–497. https://doi.org/10.3390/encyclopedia2010031

[4] Madhavi, M. K., Subramanian, M., & Shenbagavalli, R. (2020). Blockchain technology: The new avatar in the world of banking and finance. International Journal of Advanced Research in Engineering and Technology, 11(12), 964–968. https://doi.org/10.34218/IJARET.11.12.2020.096

[5] Sinfield, V., Ajmani, A., & McShane, W. (2024). Strategic roadmapping to accelerate and risk-mitigate enabling innovations: A generalizable method and a case illustration for marine renewable energy. Technological Forecasting and Social Change, 201, 123761. https://doi.org/10.1016/j.techfore.2024.123761

[6] Issa, T., & Isaias, P. (2015). Usability and human computer interaction (HCI). In Sustainable design: HCI, usability and environmental concerns (pp. 19–36). Springer. https://doi.org/10.1007/978-1-4471-6753-2_2

[7] Barta, S., Ibáñez-Sánchez, S., Orús, C., & Flavián, C. (2024). Avatar creation in the metaverse: A focus on event expectations. Computers in Human Behavior, 156, 108192. https://doi.org/10.1016/j.chb.2024.108192

[8] Garcia, M. L. (1997). Introduction to technology roadmapping: The semiconductor industry's technology roadmapping process. USDOE Office of Financial Management and Controller.

[9] Sadeghi, M. T., Sobhani, F. M., & Ghatari, A. R. (2018). Representing a model to measure absorbency of information technology in small and medium-sized enterprises. Journal of

Information Systems and Telecommunication (JIST), 20(5), 242–251.

[10] Shneiderman, B., Cohen, M., Jacobs, S., Plaisant, C., Diakopoulos, N., & Elmqvist, N. (2017). Designing the user interface (6th ed.). Pearson International. https://elibrary.pearson.de/book/99.150005/9781292153926

[11] Hu, B. (2024). Developing an avatar-based framework for unified client identification in banking systems using generative AI and graph neural networks. Innovation in Science and Technology, 3(4), 1–34. https://doi.org/10.56397/IST.2024.07.01

[12] Oyetunji, D. J. (Unpublished). The role of artificial intelligence and machine learning in enhancing customer experience in Nigeria digital banks.

[13] Szolin, K., Kuss, D., Nuyens, F., & Griffiths, M. (2022). Gaming disorder: A systematic review exploring the user-avatar relationship in videogames. Computers in Human Behavior, 128, 107124. https://doi.org/10.1016/j.chb.2021.107124

[14] Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2022). An emerging theory of avatar marketing. Journal of Marketing, 86(1), 67–90. https://doi.org/10.1177/0022242921996646

[15] Sousa, K. S., & Furtado, E. (Unpublished). RUPi – A unified process that integrates.

[16] Aguirre-Rodriguez, A., Bóveda-Lambie, A. M., & Miniard, P. W. (2015). The impact of consumer avatars in internet retailing on self-congruity with brands. Marketing Letters, 26(4), 631–641. https://doi.org/10.1007/s11002-014-9296-z

[17] Gammoh, F., Jiménez, F. R., & Wergin, R. (2018). Consumer attitudes toward human-like avatars in advertisements: The effect of category knowledge and imagery. International Journal of Electronic Commerce, 22(3), 325–348. https://doi.org/10.1080/10864415.2018.1462939

[18] Kerr, C., & Phaal, R. (2022). Roadmapping and roadmaps: Definition and underpinning concepts. IEEE Transactions on Engineering Management, 69(1), 6–16. https://doi.org/10.1109/TEM.2021.3096012

[19] Procter, L. (2021). I am/we are: Exploring the online self-avatar relationship. Journal of Communication Inquiry, 45(1), 45–64. https://doi.org/10.1177/0196859920961041

[20] Xie, L., & Lei, S. (2022). The nonlinear effect of service robot anthropomorphism on customers' usage intention: A privacy calculus perspective. International Journal of Hospitality Management, 107, 103312. https://doi.org/10.1016/j.ijhm.2022.103312

[21] Wang, X., Butt, A. H., Zhang, Q., Shafique, N., & Ahmad, H. (2021). "Celebrity avatar" feasting on in-game items: A gamers' play arena. SAGE Open, 11(2), 215824402110157. https://doi.org/10.1177/21582440211015716

[22] Foster, K., McLelland, M. A., & Wallace, L. K. (2022). Brand avatars: Impact of social interaction on consumer-brand relationships. Journal of Research in Industrial Medicine, 16(2), 237–258. https://doi.org/10.1108/JRIM-01-2020-0007

[23] Triantoro, T., Gopal, R., Benbunan-Fich, R., & Lang, G. (2020). Personality and games: Enhancing online surveys through gamification. Information Technology & Management, 21(3), 169–178. https://doi.org/10.1007/s10799-020-00314-4

[24] Gonzales-Chavez, M. A., & Vila-Lopez, N. (2021). Designing the best avatar to reach millennials: Gender differences in a restaurant choice. Industrial Management & Data Systems, 121(6), 1216–1236. https://doi.org/10.1108/IMDS-03-2020-0156

[25] Takano, M., & Taka, F. (2022). Fancy avatar identification and behaviors in the virtual world: Preceding avatar customization and succeeding communication. Computers in Human Behavior Reports, 6, 100176. https://doi.org/10.1016/j.chbr.2022.100176

[26] Kramer, M. (2018). Best practices in systems development lifecycle: An analysis based on the waterfall model. Review of Business & Finance Studies, 9(1), 77–84. https://ssrn.com/abstract=3131958

[27] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). Human-computer interaction. Addison-Wesley.

[28] Li, Y., Dai, G., Chen, G., Liu, J., Li, P., & Ip, H. H. (2025). Avatar-mediated communication in collaborative virtual environments: A study on users' attention allocation and perception of social interactions. Computers in Human Behavior, 108598. https://doi.org/10.1016/j.chb.2025.108598

[29] Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. Journal of Marketing, 70(4), 19–36. https://doi.org/10.1509/jmkg.70.4.019

[30] Rasouli, M. R. (2012). The role of media in institutionalizing the culture of electricity consumption management. Journal of Communication Culture Science Promotion, 2(5), 93–105. [In Persian]

[31] Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. IEEE Transactions on Engineering Management, 48(2), 132–143. https://doi.org/10.1109/17.922473

[32] Hffer, J. A., George, J. F., & Valacich, J. S. (2005). Modern systems analysis and design (4th ed.). Prentice Hall.

[33] Hugues, O., Fuchs, P., & Nannipieri, O. (2011). New augmented reality taxonomy: Technologies and features of augmented environment. In B. Furht (Ed.), Handbook of augmented reality (pp. 47–63). Springer. https://doi.org/10.1007/978-1-4614-0064-6_2

[34] Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, E77-D(12), 1321–1329.

[35] Carey, J., Galletta, D., Kim, J., Te'eni, D., Wildermuth, B., & Zhang, P. (2004). The role of HCI in IS curricula: A call to action. Communications of the AIS, 13(23), 357–379. https://doi.org/10.17705/1CAIS.01323

[36] Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., et al. (1992). ACM SIHCHI curricula for human-computer interaction. Association for Computing Machinery. https://doi.org/10.1145/2594128

[37] Zhang, P., Benbasat, I., Carey, J., Davis, F., Galletta, D., & Strong, D. (2002). Human-computer interaction research in the MIS discipline. Communications of the AIS, 9(20), 334–355. https://doi.org/10.17705/1CAIS.00920

[38] Wright, P., Blythe, M., & McCarthy, J. (2006). User experience and the idea of design in HCI. In S. W. Gilroy & M. D. Harrison (Eds.), Interactive systems. Design, specification, and verification. DSV-IS 2005. Lecture Notes in Computer Science (Vol. 3941, p. 1). Springer. https://doi.org/10.1007/11752707_1

[39] Lam, C.-M. (2025). Building ethical virtual classrooms: Confucian perspectives on avatars and VR. Computers & Education: X Reality, 6, 100092. https://doi.org/10.1016/j.cexr.2024.100092
[40] Andrew, P. S., & Palmer, J. D. (1990). Software systems engineering. John Wiley & Sons.
[41] Torabi, M. A., Hasangholipour Yasori, T., & Zare, M. J. (2023). Conceptualization and theorizing avatar marketing in Iran. Journal of Business Management, 15(2), 185–216. https://doi.org/10.22059/jibm.2022.344989.4399
[42] Hosseini, M., & Kiadehi, E. F. (2014). Internet banking, cloud computing: Opportunities, threats. Journal of Information Systems and Telecommunication (JIST), 6, 1–10. https://doi.org/10.7508/jist.2014.02.002
[43] Silva, M. J. B., Delfino, L. O. R., Cerqueira, K. A., & Campos, P. O. (2022). Avatar marketing: A study on the engagement and authenticity of virtual influencers on Instagram. Social Network Analysis and Mining, 12(1), 130. https://doi.org/10.1007/s13278-022-00966-w
[44] Lin, Y.-T., Doong, H.-S., & Eisingerich, A. B. (2021). Avatar design of virtual salespeople: Mitigation of recommendation conflicts. Journal of Service Research, 24(1), 141–159. https://doi.org/10.1177/1094670520964872
[45] Yan, J., Ma, T., & Nakamori, Y. (2011). Exploring the triple helix of academia-industry-government for supporting roadmapping in academia. International Journal of Management and Decision Making, 11(3), 249–267. https://doi.org/10.1504/IJMDM.2011.040702
[46] McDowell, W., & Eames, M. (2006). Forecasts, scenarios, vision, backcasts and roadmaps to the hydrogen economy: A review of the hydrogen futures literature. Energy Policy, 34(11), 1236–1250. https://doi.org/10.1016/j.enpol.2005.12.006
[47] Jones, L. E., Hancock, T., Kazandjian, B., & Voorhees, C. M. (2022). Engaging the avatar: The effects of authenticity signals during chat-based service recoveries. Journal of Business Research, 144, 703–716. https://doi.org/10.1016/j.jbusres.2022.01.012
[48] Ghafarzadegan, M., & Peymankhah, S. (2007). A comparative analysis of common approaches in roadmap development in technology management. In The 5th International Conference on Management, Tehran, 2007. https://civilica.com/doc/43510
[49] Carroll, J. M., & Rosson, M. B. (2003). Design rationale as theory. In HCI models, theories, and frameworks: Toward a multidisciplinary science (pp. 431–461). Elsevier. https://doi.org/10.1016/B978-155860808-5/50015-0
[50] Ziaeipour, E., Ghotri, A. R., & Taghizadeh, A. (2023). Software-defined networking adoption model: Dimensions and determinants. Journal of Information Systems and Telecommunication (JIST), 44, 368–382. https://doi.org/10.61186/jist.40088.11.44.368
[51] Paul, S., Mohanty, S., & Sengupta, R. (2022). The role of social virtual world in increasing psychological resilience during the on-going COVID-19 pandemic. Computers in Human Behavior, 127, 107036. https://doi.org/10.1016/j.chb.2021.107036
[52] Hanus, M. D., & Fox, J. (2015). Persuasive avatars: The effects of customizing a virtual salesperson's appearance on brand liking and purchase intentions. International Journal of Human-Computer Studies, 84, 33–40. https://doi.org/10.1016/j.ijhcs.2015.07.004

Journal of Information Systems and Telecommunication
Vol.13, No.4, October-December 2025, 316-325
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Optimally DBS Placement In 6G Communication Networks Using Improved Gray Wolf Optimization Algorithm to Enhance Network Energy Efficiency

Hussein Shakir Diwan Al-Khulaifawi [1], Mahdi Nangir [1*]

[1].Department of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran.

Received: 17 May 2025/ Revised: 04 Sep 2025/ Accepted: 25 Nov 2025

Abstract

The transition to sixth-generation (6G) networks demands highly energy-efficient solutions for large-scale IoT services. Drone Base Stations (DBSs) offer flexible coverage, but their three-dimensional placement must be optimized to reduce both transmission and hovering energy. This paper, model DBS deployment as a power-minimization problem and introduce an Improved Grey Wolf Optimization (IGWO) algorithm that integrates adaptive control parameters, exponential weighting of leader contributions (alpha/beta/delta), and a dynamic control structure that progressively favors elite solutions. This design improves search efficiency in high-dimensional, nonlinear spaces and reduces the risk of premature convergence. Extensive MATLAB simulations across multiple propagation environments demonstrate that IGWO achieves lower network power consumption and faster convergence compared to standard metaheuristics, while preserving coverage and connectivity. Specifically, the simulation results demonstrate that the proposed method achieves a remarkable superiority over other optimization algorithms, showing more than a 2% improvement compared to the best among them the standard GWO algorithm—thereby confirming its effectiveness and efficiency in low-power network scenarios.

Keywords: 6G Communication Networks; Drone Base Stations (DBSs); Internet of Things (IoT); Improved Gray Wolf Optimization (IGWO); Energy Efficiency.

1- Introduction

The emergence of 6G communication networks is a significant step forward in wireless technology, which provides ultra-high capacity, ultra-reliable low-latency communication and these technological advancements have been accompanied by the use of DBSs which have offered a practical means of addressing the growing and geographically dispersed needs for wireless services, especially in areas where conventional tower-based networks are constrained or unable to adjust [1-3]. Incorrect positioning may lead to signal losses, higher energy needs, and degraded network capabilities, particularly in areas with numerous constructions. Thus, it is necessary to implement a thoughtful and organized strategy for the three-dimensional distribution of DBSs in order to optimize the potential capabilities of 6G networks [4,5]. Conventional optimization works often fail to provide globally optimal solutions because of the intricate, ever-changing, and multi-layered nature of DBS placement. Conversely, metaheuristic algorithms inspired by the dynamics of nature and society are gaining increasing attention for their reliability and efficiency in the face of the intricacies of optimization problems [6].

Authors in [7] propose an optimized method for DBS placement using the Marine Predators Algorithm (MPA), which is good at avoiding local optima. Through simulation, their approach outperforms previous techniques, with an average path loss of 56.13 dB, which significantly improves path loss mitigation and user access. The work in [8] describes the quasi-opposition-based lemurs optimizer (QOBLO), a new method of using lemur foraging strategies with quasi-opposition learning to optimally deploy DBS in NG-I. QOBLO outperforms other swarm methods, as per thorough simulations and statistical analysis, markedly increasing connectivity, coverage, and energy efficiency, and providing a strong scalable solution for 6G network problems. In [9], researchers present a two-layer optimizer using a pre-trained VGG-19 model and micro-swarms to optimize network performance by means of non-orthogonal multiple access. It is demonstrated that after statistical testing, the method obtains a 98% accuracy of results

✉ **Mahdi Nangir**
nangir@tabrizu.ac.ir

when compared to Cuckoo Search, Grey Wolf, and Particle Swarm Optimization.

In [10], an analysis of a wireless architecture where aerial and terrestrial base stations serve respective users is carried out, with emphasis on how ABS height and transmit power alter rates for downlink and uplink communication. The results show that optimal ABS configurations are often at the maximum or minimum extremity, and factors like user distance affect performance. Based on [11] where a multi-UAV communication setting is addressed, the authors formulate a multi-objective optimization problem, CUEMOP, to pursue improved coverage and energy saving. The authors propose the Improved Multi-objective Grey Wolf Optimizer (ImMOGWO) which includes the clustering, hybrid initialization techniques, and innovations related to the Levy flight algorithms. It is demonstrated that trial simulations show that ImMOGWO has better efficiency and solution quality than benchmark algorithms.

In [12], researchers conduct systematic mapping analysis of 3D placement in communication systems with UAVs, analyzing goals of optimization, system models, and solution techniques. The study indicates that there is a focus on optimizing data rate, power and coverage using large scale fading models, heuristic algorithms dominate, and there is a lack of significant work on outage probability, cost, and quality of experience and spectrum optimization. In [13], the researchers propose a Mixed-Integer Non-Linear Programming method for coordinating DBS location optimization and minimization of their number, using a modified PSO algorithm that begins with K-means-based initialization. A unique communication protocol is established and simulation results prove the approach offers low packet loss, minimized latency, and extensive user coverage across various environments.

In [14], the DBS placement problem is addressed using P-median optimization; fuzzy clustering is used to generate candidate positions and a bisection algorithm is used to determine the optimum number of DBSs. The optimization solution yields better results than rival approaches, especially when the clustering parameters are adjusted with high precision. The authors in [15] perform an assessment of a variety of existing swarm intelligence algorithms including Cuckoo Search (CS), Elephant Herd Optimization (EHO), Grey Wolf Optimization (GWO), Monarch Butterfly Optimization (MBO), Salp Swarm Algorithm (SSA), and Particle Swarm. They examine how well and productively these algorithms solve a specified problem, carrying out tests in various scenarios. To systemically assess the algorithms, the authors use the Friedman and Wilcoxon tests. Through the use of these tests, the study creates a foundation for performance disparities evaluation and identifies the most effective swarm intelligence methods for dealing with the problem.

This study employs an Improved Grey Wolf Optimization (IGWO) algorithm for the optimal placement of drone base stations (DBSs) within 6G cellular networks, with the primary objective of minimizing network power consumption. Owing to its high capability in navigating complex, high-dimensional search spaces, the IGWO algorithm rapidly converges toward optimal solutions. This characteristic proves particularly advantageous for the placement of DBSs, as it significantly reduces computational time while achieving near-optimal configurations. Furthermore, the IGWO algorithm maintains a balance between local exploitation and global exploration. This adaptive balance mitigates the risk of entrapment in local optima and facilitates the discovery of more globally efficient placement strategies for the DBSs.

The key contributions of this study include:

An optimization framework is formulated to minimize the average power consumption of ground users by strategically deploying DBSs. Given the high-dimensional and nonlinear nature of the problem space, the Improved Gray Wolf Optimization (IGWO) algorithm, rooted in swarm intelligence, is utilized. The algorithm adaptively maintains a dynamic balance between exploration and exploitation, thereby reducing the likelihood of premature convergence and enhancing the algorithm's ability to approximate the global optimum effectively.

A dynamic weighting mechanism is introduced to reinforce gradual exploitation. In this mechanism, the weights assigned to the alpha, beta, and delta wolves are updated iteratively using exponential functions. As the iterations progress, increased emphasis is placed on the alpha wolf's position, thereby enhancing the algorithm's ability to exploit the most promising solution discovered thus far and leading to more precise convergence behavior.

A dynamic control structure is also developed to gradually intensify the influence of elite solutions over time. Unlike conventional approaches that uniformly aggregate the guidance from all reference wolves, this method employs a targeted weighting strategy. This allows the search process to be progressively steered toward more reliable regions of the solution space. Such structural modification in information aggregation significantly enhances the algorithm's performance in complex and dynamic wireless communication environments.

The efficacy of collective intelligence-based techniques for identifying the optimal position of drone base stations has been assessed through extensive simulations. The superiority of the suggested approach in reducing average power consumption has been demonstrated by a comparative analysis conducted under various environment circumstances, search agent counts, and user densities.

The remainder of this paper is organized as follows: The suggested methodology is presented in Section 2. The simulation settings and performance evaluation processes

are described in Section 3, and the paper's conclusion and future research prospects are outlined in Section 4.

2- Proposed Method

The primary objective of this study is to propose an effective methodology for the optimal placement of drone base stations (DBSs) within 6G cellular networks, aiming to minimize overall network power consumption. To achieve this, an Improved Gray Wolf Optimization (IGWO) algorithm is employed. The IGWO algorithm maintains an effective trade-off between local exploitation and global exploration. This balance significantly contributes to avoiding local optima and facilitates the discovery of more efficient deployment strategies for DBSs. Owing to its high flexibility, IGWO exhibits strong adaptability to dynamic network environments and variable conditions—such as fluctuating user densities and evolving network demands—allowing it to consistently determine optimal base station locations in real time. Moreover, compared to conventional metaheuristic approaches, the IGWO algorithm demonstrates greater stability in producing reliable solutions and shows robust performance under the diverse challenges inherent in 6G communication networks.

2-1- System Model

This section outlines the system model used for evaluating the service provisioning capabilities of DBSs to Internet of Things (IoT) devices. The conceptual system architecture is illustrated in Figure 1. In the presented structure $S_{device}=\{1,2,...,s\}$ denotes the set of IoT devices randomly distributed within a two-dimensional area, and $K_{DBS}=\{1,2,...,k\}$ represents the set of DBSs deployed to serve these devices. Each DBS hovers above the device layer.

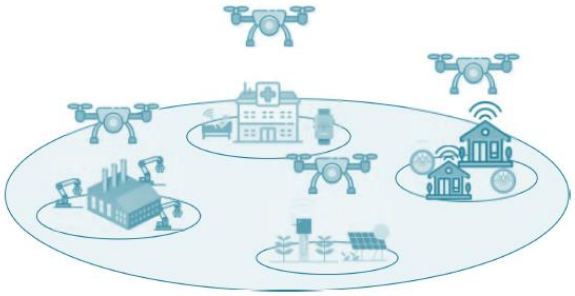

Fig.1. Conceptual System Model

Traditional channel models are insufficient for accurately simulating air-to-ground (AtG) communication due to the altitude variability of DBSs. Instead, two primary link types are considered for modeling the relationship between DBSs and IoT devices: Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) connections.

2-2- Air-to-Ground Propagation Model

The probability of establishing a Line-of-Sight (LoS) link between the *k*-th DBS and the *s*-th IoT device is given by the following expression:

$$P(h_k.d_{k,s}) = \frac{1}{1+\alpha\exp\left[-\beta\left(\arctan\left(\frac{h_k}{d_{k.s}}\right)-\alpha\right)\right]}, \quad (1)$$

where α and β are environment-dependent parameters, h_k denotes the k-th DBS altitude, and $d_{k.s}$ is the horizontal distance between the DBS and the IoT device, defined as:

$$d_{k.s} = \sqrt{(x_k - x_s)^2 + (y_k - y_s)^2}. \quad (2)$$

Here, $(y_k.x_k)$ and $(y_s.x_s)$ represent the 2D coordinates of the k-th DBS and the IoT device, respectively.
Using the LoS and NLoS probabilities, the path loss can be modeled as:

$$\text{PL}\,(h_k.d_s) = 20\log\left(\sqrt{h_k^2 + d_{k.s}^2}\right) + AP(h_k.d_{k.s}) + B \quad (3)$$

where:

$$A = \eta_{LoS} - \eta_{NLoS}, \quad (4)$$

$$B = 20\log\left(\frac{4\pi f_c}{c}\right) + \eta_{NLoS}. \quad (5)$$

In these equations:

- η represents the mean additional path loss;
- A is the differential loss between LoS and NLoS conditions;
- f_c is the carrier frequency (in Hz);
- c denotes the speed of light.

2-3- Objective Function for Optimal DBS Placement

The central goal of this research is to determine optimal placements for the DBSs that minimize the total power consumption of the network. This objective is formulated as an optimization problem and is addressed using the proposed IGWO metaheuristic algorithm. Given that the objective function plays a pivotal role in the design of any metaheuristic optimization strategy, it is formally defined in this section to guide the optimization process effectively.

2-3-1 Minimizing Network Power Consumption

The objective of this section is to present a comprehensive model for calculating the total power consumption of the network, incorporating the energy required for electronic processing, average data transmission time, path loss, and

other real-world parameters. To this end, the transmitter's power consumption can be considered to comprise two components: a fixed amount of electronic energy required for processing, and the transmission energy component, which depends on the path loss. Consequently, the average power consumption for communication between the *s*-th user device and the *k*-th Drone Base Station (DBS) can be expressed as:

$$P_{cons_{ave}}(h_k.d_s) = \left(E_{elec} + \varepsilon_{amp-tx} \cdot \mathrm{PL}\,(h_k.d_s)\right) \cdot \frac{K}{T_{Ave}} \quad (6)$$

where, E_{elec} is the energy required for processing each bit electronically and is measured in joules (J), ε_{amp-tx} represents the amplifier efficiency needed to compensate for the path loss during transmission and is also expressed in joules (J), K is the number of bits transmitted, and T_{Ave} denotes the average data transmission time in seconds (s).
It is important to note that ε_{amp-tx} quantifies the energy consumed per bit to overcome the attenuation in the signal path and is determined based on the path loss intensity PL. Accordingly, the total average energy consumed across the network—borne by the devices—can be minimized by optimizing the placement of DBSs. Assuming that *s* indexes the devices and *k* indexes the DBSs, and that each device connects to the nearest DBS, the optimization problem can be formulated as follows:

$$\begin{aligned} &\underset{\{x.y.h\}}{\text{minimize}} \frac{\sum_{k=1}^{K}\sum_{S=1}^{S} P_{cons_{ave}}(h_k.d_s)}{S} \\ &\text{subject to: } \mathcal{C}1: x_{min} \le x_D^k \le x_{max}.\forall k \\ &\qquad\qquad \mathcal{C}2: y_{min} \le y_D^k \le y_{max}.\forall k \\ &\qquad\qquad \mathcal{C}3: h_{min} \le h_D^k \le h_{max}.\forall k \end{aligned} \quad (7)$$

Here, x, y and h represent the 3D spatial coordinates of every DBS, while x_{min}/x_{max} , y_{min}/y_{max} and h_{min}/h_{max} define the boundaries of the deployment region.

2-4- Optimal Placement of Drone Base Stations Using the Improved Grey Wolf Optimization (IGWO) Algorithm

In this study, the Improved Grey Wolf Optimization (IGWO) algorithm is employed to determine the optimal positioning of drone base stations (DBSs), with the aim of minimizing the power consumption of Internet of Things (IoT) user devices as defined by the objective functions.
The Grey Wolf Optimizer (GWO) is a nature-inspired metaheuristic algorithm that mimics the social hierarchy and hunting behavior of grey wolves in the wild. It is particularly effective for solving complex optimization problems. In this algorithm, a population of "wolves" represents candidate solutions in the search space. The optimization process begins with evaluating each wolf's position and identifying the top solutions, referred to as the alpha, beta, and delta wolves. The leaders direct the other wolves as they iteratively update their positions based on these until certain termination conditions are satisfied, like a convergence threshold or maximum number of iterations. The final position of the alpha wolf is considered the optimal solution. Due to its simplicity and efficiency, GWO has attracted considerable interest in both academic and industrial optimization tasks. The main procedural steps of the IGWO algorithm are as follows:

Step 1: Initialization of Parameters

Initially, key factors including the number of wolves (*N*), number of variables (problem dimensions, *D*), number of iterations (*T*), and the control vector (*a*) are defined. The control vector *a*, which linearly decreases from 2 to 0 over the iterations, balances the exploration and exploitation phases of the algorithm. The decrease in the value of *a* enables the algorithm to initially conduct a wide-ranging search (exploration), and later to focus on the best regions (exploitation). This vector is defined as follows:

$$\vec{a}(t) = 2 - \frac{2t}{T}\,. \quad (8)$$

Step 2: Population

After initializing the parameters, a population of grey wolves—representing potential solutions—is randomly generated within the search space. The initial position of each wolf is determined as follows:

$$X_{i.j} = rand(0.1)\ \cdot \left(ub_j - lb_j\right) + \ lb_j, \quad (9)$$

where:

- $X_{i.j}$ is the j-th variable for the i-th wolf,
- ub_j and lb_j are the upper and lower bounds of the j-th variable, and
- $rand(0.1)$ is a uniformly distributed random number between 0 and 1.

Step 3: Evaluation of Objective Function

Each wolf's position is evaluated using the objective function:

$$f_i = f(X_i)\,. \quad (10)$$

Step 4: Social Hierarchy Assignment

This step reflects the social behavior of grey wolves in nature, where the leader directs the hunting group. In this stage, based on the fitness values obtained in the previous step, the wolves are divided into four categories: Alpha, Beta, Delta, and Omega. The Alpha, Beta, and Delta

wolves act as the leaders of the hunt, while the Omega wolves follow the leaders. Therefore, the wolf with the best fitness value is selected as the Alpha wolf X_α, and the wolves with the second and third best fitness values are selected as the Beta X_β and Delta X_δ wolves, respectively. The remaining wolves are classified as Omega wolves X_ω.

Step 5: Modeling the Hunting Behavior and Position Update of Grey Wolves

The alpha, beta, and delta wolves' positions are used to update the wolves' positions at this point. To facilitate this, the control vectors A and C are defined as follows:

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r}_1 - \vec{a} \ , \tag{11}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \ . \tag{12}$$

Here, $\vec{r}_1$ and $\vec{r}_2$ are randomly generated vectors of $[0,1]^D$. Next, the relative distance and estimated positions with respect to the alpha, beta, and delta wolves are calculated using the following relations:

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right| \ \Rightarrow \ \vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \ , \tag{13}$$

$$\vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right| \ \Rightarrow \ \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \ , \tag{14}$$

$$\vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right| \ \Rightarrow \ \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta . \tag{15}$$

Finally, the wolves' final positions are updated according to following equation:

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \ . \tag{16}$$

As evident in the Eq. (16) above, the influence of the alpha, beta, and delta wolves is equally weighted in determining the optimal position. However, since the alpha wolf typically represents a better solution than the beta, and the beta better than the delta, assigning adaptive weights to each of their contributions can lead to more effective convergence toward the global optimum.

In the proposed IGWO algorithm, the initial weights assigned to the alpha, beta, and delta wolves are equal, which supports the exploration phase by allowing the algorithm to broadly search the solution space. To enhance the exploitation phase over time, these weights are adaptively adjusted throughout the iterations. Specifically, the weight of the alpha wolf gradually increases, directing more focus on the region near the current best solution, while the weights of the beta and delta wolves decrease, thus reducing their influence. This adaptive weighting strategy ensures a balanced transition from exploration to exploitation, allowing the algorithm to converge effectively to an optimal or near-optimal solution by the end of the search process. The updated position equation with adaptive weighting becomes:

$$\vec{X}(t+1) = \frac{w_\alpha(t)\vec{X}_1(t) + w_\beta(t)\vec{X}_2(t) + w_\delta(t)\vec{X}_3(t)}{3}, \tag{17}$$

where w_α, w_β, and w_δ are the time-dependent adaptive weights for the alpha, beta, and delta wolves, respectively, defined as:

$$w_\alpha(t) = w_{\alpha\_ini} + (1 - w_{\alpha\_ini}) \cdot (1 - e^{-\frac{5t}{T}}) \ , \tag{18}$$

$$w_\beta(t) = w_{\beta\_ini} \cdot e^{-\frac{2t}{T}}, \tag{19}$$

$$w_\delta(t) = w_{\delta\_ini} \cdot e^{-\frac{4t}{T}} \ . \tag{20}$$

Here, w_{α_ini}, w_{β_ini}, and w_{δ_ini} represent the initial weights for updating the positions of the Alpha, Beta, and Delta wolves. At the beginning of the algorithm, these initial weights are considered equal, following the standard GWO procedure. However, as the algorithm progresses, the weights $w_\alpha(t)$, $w_\beta(t)$ and $w_\delta(t)$ change over time to enhance the algorithm's exploitation capability. Specifically, the weight for the Alpha wolf, which has the best position, increases exponentially. Meanwhile, the weights for the Beta and Delta wolves gradually decrease as the algorithm advances. The exponential coefficients are tuned such that the slope of the decrease in $w_\beta(t)$ is less steep than the slope of the decrease in $w_\delta(t)$.

Step 6: Iterative Execution Until Convergence Criterion Is Met

Steps 3 through 5 are executed iteratively until the stopping condition—typically the maximum number of iterations—is satisfied. Upon termination, the final position of the alpha wolf $\vec{X}_\alpha$, representing the optimal coordinates of the drone base stations, is returned as the best solution obtained by the Improved Grey Wolf Optimization algorithm.

The pseudo-code of proposed method is illustrated in Algorithm 1.

2-4-1Computational complexity of the proposed IGWO

The computational complexity of the proposed IGWO algorithm is determined by the population size N, search space dimension D, and maximum number of iterations T. In each iteration, the algorithm evaluates the objective function for all search agents and updates their positions, leading to a total complexity of **O(N×D×T)**. The additional adaptive weighting and parameter control mechanisms require only simple arithmetic operations, resulting in negligible extra cost. Hence, the proposed IGWO maintains a linear computational complexity similar to the standard GWO.

Algorithm 1: Improved Grey Wolf Optimization (IGWO)

Input:
- Objective function f(x)

- Search space dimension D
- Population size N
- Maximum iterations T

Output:

- Best solution x_α (corresponding to minimum power consumption)

1: Initialize positions of N grey wolves {xi}, i = 1,...,N randomly within bounds
2: Evaluate fitness f(xi) for all wolves
3: Identify three best solutions: α (best), β (second best), δ (third best)
4: Set iteration counter t = 1
5: while (t ≤ T) do
6: Update control parameter a(t) using adaptive rule
7: For each wolf i = 1 to N do
8: Compute coefficient vectors $A = 2ar_1 - a,\ C = 2r_2$
9: Calculate distances to leaders:
$D_\alpha = |C_1 \cdot X_\alpha - X_i|$
$D_\beta = |C_2 \cdot X_\beta - X_i|$
$D_\delta = |C_3 \cdot X_\delta - X_i|$
10: Update candidate position:
$X_1 = X_\alpha - A_1 \cdot D_\alpha$
$X_2 = X_\beta - A_1 \cdot D_\beta$
$X_3 = X_\delta - A_1 \cdot D_\delta$
11: Update position:
$X_i(t+1) = \frac{w_\alpha(t)X_1 + w_\beta(t)X_2 + w_\delta(t)X_3}{3}$
(where $w_\alpha, w_\beta, w_\delta$ are dynamic weights)
12: end for
13: Evaluate new fitness values f(xi)
14: Update α, β, δ if better solutions are found
15: t = t + 1
16: end while
17: Return x_α as the best solution

3- Performance Evaluation

The effectiveness of the suggested approach in determining the best location for drone base stations (DBSs) under varied parameter settings is thoroughly evaluated in this section. The proposed approach is assessed through numerical results derived from extensive software-based simulations. The conducted experiments investigate the impact of several key factors, including the number of users, the number of search agents (population size), the number of iterations (generations), and different propagation environments—namely suburban, urban, dense urban, and high-rise urban—on path loss and power consumption in the Improved Grey Wolf Optimization (IGWO) algorithm. Table 1 provides a summary of the different parameters related to various environments. It should be noted that the parameters related to environmental modeling, simulation parameters, and parameters related to optimization algorithms are inspired by reference [15], which deals with the optimal location of DBSs using meta-heuristic algorithms in telecommunication networks. Also, the simulations were carried out using MATLAB 2023a. In addition all simulation results presented in the paper are obtained by averaging over 50 independent runs of the proposed algorithm to account for its stochastic nature and ensure reliability.

Table 1: Propagation Parameters in Different Environments

Environment	α	β	η_{Los}	ηN_{Los}
Urban	9.61	0.16	1	20
Suburban	4.88	0.43	0.1	21
Dense Urban	12.08	0.11	1.6	23
High-raise Urban	27.23	0.08	2.3	34

3-1- Path Loss Evaluation

In this section, the effect of four different factors on path loss is investigated: population size in optimization algorithms, maximum number of iterations, type of propagation environment, and the number of users.

Experiment 1: Impact of Propagation Environment on Path Loss

This experiment evaluates the impact of different propagation environments on the average path loss. For this purpose, the number of users is set to 20, the maximum number of iterations is 100, and the population size (number of search agents) is 25. The detailed parameters of this experiment are represented in Table 2.

Table 2: Experiment 1 Simulation Parameters

Parameter	***Value***
Number of Users	20
Maximum number of iterations	100
Various Environments	Urban- suburban-dense urban, high-rise urban
Number of Search Agent	25

Figure 2 illustrates the effect of different propagation environments—suburban, urban, dense urban, and high-rise urban—on the path loss within the network. As shown

in the figure, suburban environments exhibit the lowest path loss across all optimization algorithms, while dense urban environments result in the highest path loss. Moreover, it is observed that the proposed IGWO method consistently yields lower path loss compared to other optimization techniques across all environment types, indicating the algorithm's robustness and adaptability to diverse propagation conditions.

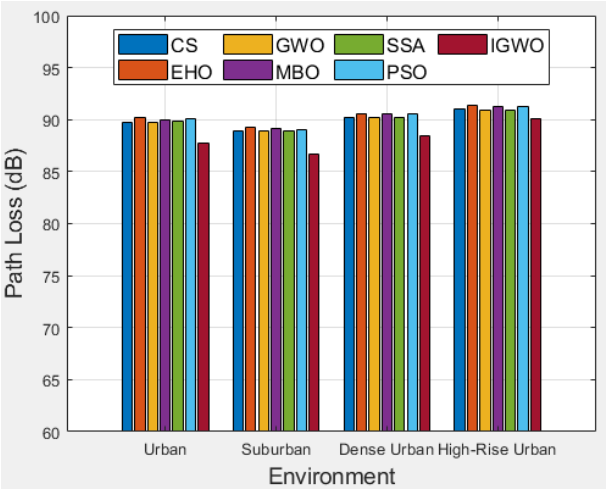

Fig.2. Effect of Path Loss in Various Environments for Different Approaches

Experiment 2: Impact of Maximum Number of Iterations on Path Loss

This investigation assesses how varying the maximum number of iterations (generations) affects the overall path loss. The simulation parameters used in this experiment are listed in Table 3.

Table 3: Experiment 2 Simulation Parameters

Parameter	***Value***
Number of Users	20
Maximum number of iterations	50, 100, 200, 500
Various Environments	Urban
Number of Search Agent	25

As illustrated in Figure 3, the path loss decreases with an increasing number of iterations for all metaheuristic algorithms. This demonstrates that increasing the number of iterations enhances the convergence and performance of optimization methods. Furthermore, the lowest recorded path loss of 86.8 dB is achieved by the proposed Improved Grey Wolf Optimization (IGWO) when the number of iterations reaches 500, highlighting the superior efficiency of the proposed algorithm in minimizing power consumption compared to alternative approaches.

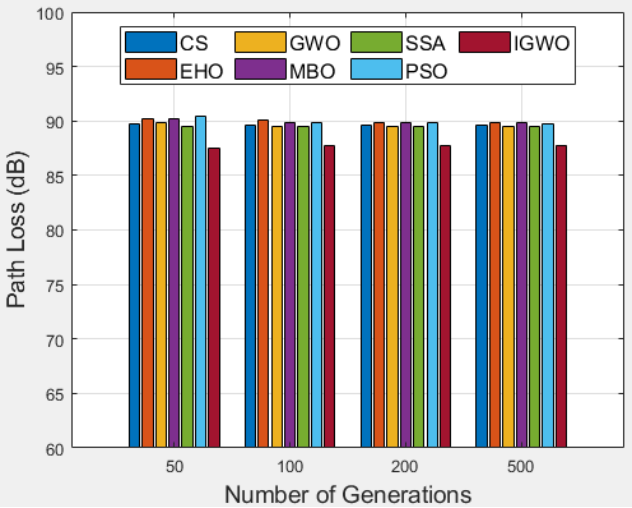

Fig.3. Effect of the Maximum Iterations on Path Loss for Different Methods

Experiment 3: Impact of the Number of Search Agents on Path Loss

The third investigation investigates the effect of varying the count of search agents on path loss for different metaheuristic algorithms, including the proposed IGWO method. The simulation considers 20 users, a maximum of 100 iterations, and an urban propagation environment. The detailed parameters of this experiment are shown in Table 4.

Table 4. Experiment 3 Simulation Parameters

Parameter	***Value***
Number of Users	20
Maximum number of iterations	100
Various Environments	Urban
Number of Search Agent	5, 25 ,50, 75, 100

Figure 4 depicts the effect of the number of search agents on the path loss. As seen in the figure, increasing the number of agents generally results in a reduction in path loss. The proposed IGWO method consistently achieves the lowest path loss across various population sizes compared to the other approaches.

Fig.4. Effect of the Number of Search Agents on Path Loss for Different Methods

Experiment 4: Impact of User Counts on Path Loss

Based on the first simulation scenario, this investigation assesses how the count of users affects the suggested method's performance. While keeping the maximum number of iterations, propagation environment, and number of search agents constant, the number of users is varied to assess its impact on the path loss. The simulation parameters are summarized in Table 5.

Table 5: Experiment 4 Simulation Parameters

Parameter	***Value***
Number of Users	10, 20, 30, 40, 50
Maximum number of iterations	100
Various Environments	urban
Number of Search Agent	25

As shown in Figure 5, the path loss increases with the number of users. Additionally, it is observed that the proposed IGWO algorithm consistently yields the lowest path loss, particularly for 10 users, further demonstrating its ability to adapt and scale across different user densities.

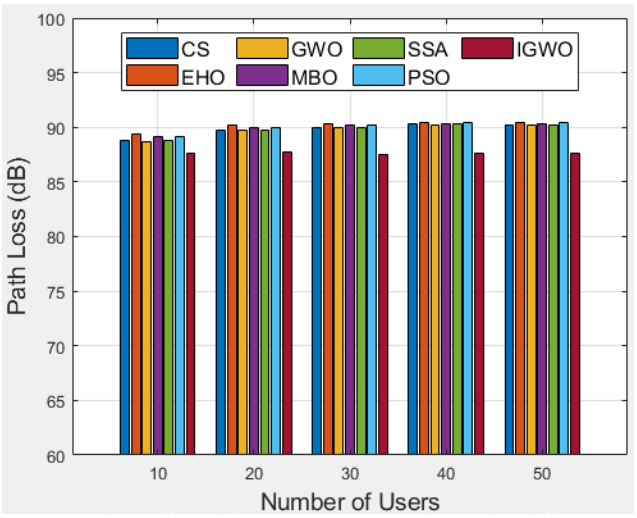

Fig.5. Effect of the User Counts on Path Loss for Different Methods

3-2- Evaluation of Average Power Consumption

This section investigates how various factors influence the average power consumption of deployed drones across four experiments. These include population size, number of iterations, propagation environment, and number of users.

Experiment 1: Effect of Propagation Environment on Power Consumption

Figure 6 presents the average power consumption for various propagation environments. The results show that the proposed IGWO algorithm consumes the least energy across all environments, with the lowest power consumption of 44 mW observed in the suburban scenario. Conversely, the highest power consumption (45.8 mW) is observed for the PSO algorithm in high-rise urban areas. Specifically, the simulation results demonstrate that the proposed method achieves a remarkable superiority over other optimization algorithms, showing more than a 2% improvement compared to the best among them—the standard GWO algorithm—thereby confirming its effectiveness and efficiency in low-power network scenarios. Furthermore, power consumption in suburban environments is generally lower for all algorithms, confirming the lower propagation loss in such environments.

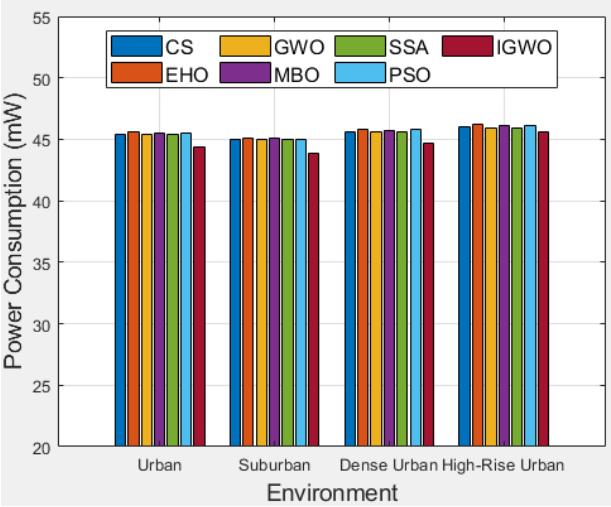

Fig.6. Effect of the Propagation Environment on Average Power Consumption for Different Methods

Experiment 2: Effect of Maximum Number of Iterations on Power Consumption

As shown in Figure 7, increasing the number of iterations significantly reduces average power consumption for all algorithms. The IGWO method achieves the minimum value of 44.6 mW at 500 iterations, while PSO shows the highest power consumption of 45.3 mW at 50 iterations. The results confirm that more iterations allow the optimization process to converge toward more energy-efficient deployments.

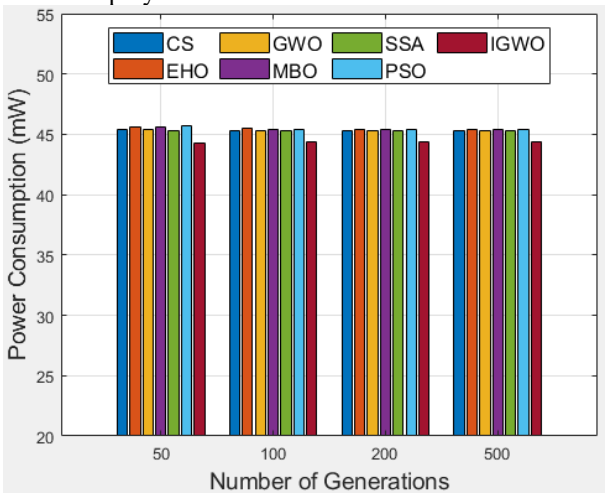

Fig.7. Effect of Maximum Iterations on Power Consumption for Different Methods

Experiment 3: Effect of Number of Search Agents on Power Consumption

In Figure 8, the results reveal that increasing the number of search agents reduces the average power consumption, as more agents improve the search space exploration and chances of finding optimal solutions. The IGWO consistently outperforms other methods, maintaining the lowest power consumption across all population sizes, demonstrating its efficient exploration and exploitation balance.

Fig.8. Effect of the Number of Search Agents on Power Consumption for Different Methods

Experiment 4: Effect of Number of Users on Power Consumption

As depicted in Figure 9, power consumption increases with the number of users, which is expected due to the higher communication and coverage demands. Nevertheless, the IGWO algorithm consistently consumes less energy than other methods across all user counts. This underscores the method's scalability and adaptability, primarily due to its dynamic balance between exploration and exploitation.

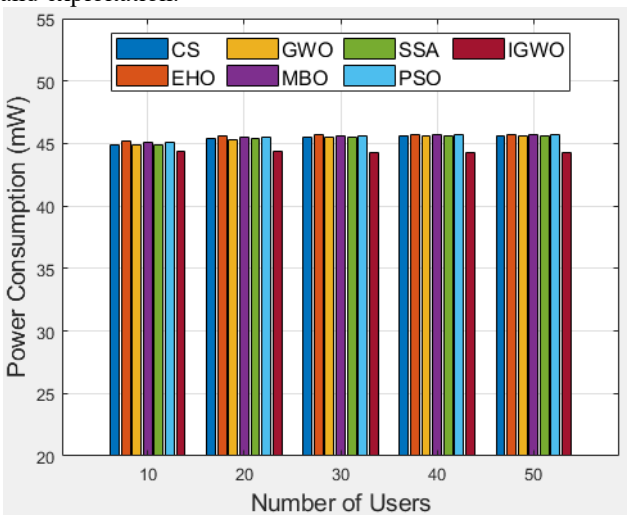

Fig.9. Effect of User Count on Average Power Consumption for Different Methods

4- Conclusion

In this study, an improved Gray Wolf Optimization (IGWO) algorithm was developed to address the urgent challenge of energy-efficient deployment of Drone Base Stations (DBSs) in 6G networks. The proposed IGWO algorithm, featuring adaptive weighting mechanisms and dynamic control structures, demonstrated enhanced capacity to explore the complex, multivariate search spaces required for effective DBS deployment. Several simulation experiments were conducted under varying conditions, including different propagation environments, population sizes, iteration counts, and user densities. The presented strategy consistently exhibited strong performance and stability across all simulations. Notably, the IGWO algorithm achieved the lowest path loss values across all propagation environments, with suburban scenarios yielding the lowest overall path loss. In the urban environment simulation, the IGWO method generated a path loss of only 86.8 dB after 500 iterations, outperforming traditional optimization methods. Analysis of average power consumption further confirmed that IGWO enables significant energy savings. The algorithm achieved an average power consumption of 44 mW in suburban areas, while also maintaining strong performance in dense and high-rise urban environments.

Increasing the number of search agents and iterations further improved the algorithm's performance, demonstrating its scalability and convergence efficiency. Even as user demand increased, the power consumption of IGWO remained systematically lower than that of all other optimization approaches. Future research could explore mobility models, examine temporal variations in user distribution, and investigate integrated optimization strategies to further enhance overall network performance.

References

[1] M. Fathi, "An Analysis of the Signal-to-Interference Ratio in UAV-based Telecommunication Networks," Journal of Information Systems and Telecommunication (JIST), vol. 1, no. 45, pp. 49, 2024.

[2] S. H. Mostafavi-Amjad, V. Solouk, and H. Kalbkhani, "Energy-efficient user pairing and power allocation for granted uplink-NOMA in UAV communication systems," Journal of Information Systems and Telecommunication (JIST), vol. 4, no. 40, pp. 312, 2022.

[3] W. Shafik, M. Ghasemzadeh, and S. M. Matinkhah, "A fast machine learning for 5G beam selection for unmanned aerial vehicle applications," Journal of Information Systems and Telecommunication (JIST), vol. 4, no. 28, pp. 262, 2020.

[4] L. Liu, A. Wang, G. Sun, and J. Li, "Multiobjective optimization for improving throughput and energy efficiency in UAV-enabled IoT," IEEE Internet of Things Journal, vol. 9, no. 20, pp. 20763-20777, 2022.

[5] H. B. Salameh, A. E. Masadeh, and G. El Refae, "Intelligent drone-base-station placement for improved revenue in B5G/6G systems under uncertain fluctuated demands," IEEE Access, vol. 10, pp. 106740-106749, 2022.

[6] Y. Luo and G. Fu, "UAV based device to device communication for 5G/6G networks using optimized deep learning models," Wireless Networks, pp. 1-15, 2023.

[7] S. Khosroabadi and H. A. Alaboodi, "Innovative Drone Base Station Placement in 6G Networks: A Marine Predators Algorithm Approach," Journal of AI and Data Mining, vol. 13, no. 2, pp. 175-182, 2025.

[8] V. Loganathan, S. Veerappan, P. Manoharan, and B. Derebew, "Optimizing Drone-Based IoT Base Stations in 6G Networks Using the Quasi-opposition-Based Lemurs Optimization Algorithm," International Journal of Computational Intelligence Systems, vol. 17, no. 1, pp. 218, 2024.

[9] H. Alsolai et al., "Optimization of Drone Base Station Location for the Next-Generation Internet-of-Things Using a Pre-Trained Deep Learning Algorithm and NOMA," Mathematics, vol. 11, no. 8, pp. 1947, 2023.

[10] M. Q. Alsudani et al., "Positioning Optimization of UAV (Drones) Base Station in Communication Networks," Malaysian Journal of Fundamental and Applied Sciences, vol. 19, no. 3, pp. 429-439, 2023.

[11] X. Zhu et al., "Multi-objective Deployment Optimization of UAVs for Energy-Efficient Wireless Coverage," IEEE Transactions on Communications, 2024.

[12] J. Carvajal-Rodríguez et al., "3D Placement Optimization in UAV-Enabled Communications: A Systematic Mapping Study," IEEE Open Journal of Vehicular Technology, 2024.

[13] F. Pasandideh et al., "An improved particle swarm optimization algorithm for UAV base station placement," Wireless Personal Communications, vol. 130, no. 2, pp. 1343-1370, 2023.

[14] M. H. Zahedi et al., "Fuzzy based efficient drone base stations (DBSs) placement in the 5G cellular network," Iranian Journal of Fuzzy Systems, vol. 17, no. 2, pp. 29-38, 2020.

[15] D. Pliatsios et al., "Drone-base-station for next-generation internet-of-things: A comparison of swarm intelligence approaches," IEEE Open Journal of Antennas and Propagation, vol. 3, pp. 32-47, 2021.

Journal of Information Systems and Telecommunication
Vol.13, No.4, October-December 2025, 326-332
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Fabric Defect Identification based on KNN and PCA Algorithms

Zahra Nouri[1], Farahnaz Mohanna[1*],Mina Boluki[1]

[1].Department of Communications Engineering, University of Sistan and Baluchestan, Zahedan, Iran

Received: 12 Apr 2025/ Revised: 09 Sep 2025/ Accepted: 26 Oct 2025

Abstract

In this study, a K-Nearest Neighbor (KNN) classifier is employed for fabric defect identification. First, directional Grey-Level Co-occurrence Matrix (GLCM) of the fabric image is computed in 0°, and 90° directions. Six intensity-based features are then extracted from these directional GLCMs. In addition, the minimum, maximum, median, and mean grey levels of the fabric image are computed. These sixteen features are combined into a single feature vector representing the fabric image. Next, Principal Component Analysis (PCA) is applied to reduce the dimensionality of the feature vector. The reduced features are then classified using the KNN classifier, categorizing each fabric image as either defective or defect-free based on training data. To localize defects, patches containing defects are segmented from the original fabric image. Features of these defect patches are extracted, reduced via PCA, and classified using KNN. Finally, each defect class is identified, and defect locations are visualized using morphological operations. The proposed method is evaluated on the comprehensive TILDA dataset, which contains 3,200) fabric images (both defective and defect-free). Experimental results demonstrate a mean average accuracy of 95.65% for fabric defect identification across classes C_1, C_2, and C_3.

Keywords: Fabric Defect Identification; Feature Extraction; KNN Classifier; PCA Algorithm.

1- Introduction

A fabric defect refers to a flaw on the surface of a fabric caused by issues in the manufacturing process. In textile quality control, fabric defect identification is crucial for maintaining product standards [1]. To date, over 70 types of fabric defects have been documented in textile manufacturing [2]. The presence of defects can reduce fabric value by 45-65% [2].

Traditionally, defect identification has relied on human visual inspection [2], which suffers from limited accuracy; typically ranging between 60-75% [3]. As a result, manual inspection is inefficient and unreliable for long-term use [1]. Therefore, automatic fabric defect identification based on computer vision techniques is increasingly important for quality control in textile production [4].

Compared to manual methods, automatic fabric defect identification offers higher accuracy, reduced costs, and increased robustness in production lines [3]. However, due to the wide variety of defect types, accurately classifying fabric defects remains a challenging task [4].

Recently, deep learning-based approaches have achieved promising results in fabric defect identification. However, their performance heavily depends on large volume of labelled data [5]. Since labelling fabric defects is time-consuming and expensive for textile factories, using supervised learning methods that do not rely on deep learning can reduce this burden. Thus, developing efficient, accurate, and lightweight methods remains a key research goal [6].

In this study, we propose an automatic method based on the K-Nearest Neighbor (KNN) classifier to identify defects in both plain and patterned fabrics. First, the directional Grey-Level Co-occurrence Matrix (GLCM) of the fabric image is computed in $0°$, and $90°$ orientations. From these GLCMs, six intensity features are extracted. Additionally, the minimum, maximum, median, and mean grey levels of the image are calculated. These sixteen features are concatenated into a single feature vector of size of 1×16. Principal Component Analysis (PCA) is then applied to reduce the dimensionality of the feature vector. The reduced features are subsequently classified using the KNN algorithm to categorize the image as either defective or defect-free.

To localize defects within defective images, defect patches are first segmented. Features from these patches are extracted, reduced using PCA, and classified via KNN.

✉ **Farahnaz Mohanna**
f_mohanna@ece,usb.ac.ir

Finally, the type and position of each defect are determined and visualized using morphological operations. The proposed method is evaluated on the comprehensive TILDA dataset, which includes 3,200 fabric images (defective and defect-free). Experimental results show a mean average accuracy of 95.65% for defect identification across 1,390 images in classes C_1 , C_2 , and C_3 of the dataset.

The main innovation of the proposed method lies in achieving high classification accuracy; 95.65%, on the TILDA dataset using only PCA and KNN, without relying on any deep or non-deep neural network models. As a result, the proposed method is suitable for real world applications in fabric defect inspection. However, it is not effective for detecting defects in a randomly patterned fabrics, as demonstrated by lower performance in such cases within the TILDA dataset.

The remainder of this paper is organized as follows: Section 2 reviews related work in fabric defect identification. Section 3 presents the proposed methodology. Section 4 reports experimental results and comparisons with existing methods. Finally, Section 5 concludes the study.

2- Past Methods

Several methods have been proposed for fabric defect identification, ranging from traditional image processing to deep learning techniques. This section reviews a selection of these approaches and their reported performance.

An unsupervised learning method [1] identified fabric defects by reconstructing image patches at multiple levels of a Gaussian pyramid. The reconstruction residuals were used for defect prediction, yielding a maximum identification accuracy of 85.20% on 128 fabric images. However, the dataset did not include all defect types. A patch-based method [2] extracted local fabric image patches, which were then labelled and fed into a pre-trained deep convolutional neural network. Defects were localized by scanning the image with the trained model. This method achieved an accuracy of 97.20% on 300 non-randomly patterned fabric images from the TILDA dataset.

A comprehensive survey [3] reviewed existing fabric defect identification algorithms and datasets, comparing identification accuracy and real-time performance. Auto-encoder networks [4] were trained using a loss function on Structural Similarity Index Measurement (SSIM). Defect identification was performed using SSIM residual maps. Identification accuracies of 92.70%, 79.80%, and 93.10% were reported for classes C_1 , C_2 , and C_3 of the TILDA dataset, with no results provided for class C_4 . A multi-task mean teacher approach [5] was presented to simultaneously identify the defect area, contour, and distance map. Supervised and consistency losses were applied to labeled and unlabeled data, respectively, resulting in a mean accuracy of 87.77% on TILDA fabric images. A weighted double low-rank decomposition technique [6] located defects by identifying homogeneous regions with high correlation. This method achieved a mean accuracy of 90.66% on 250 plain fabric images from the TILDA dataset. A method based on Elliptical Gabor filter (EGF) [7] used a particle swarm optimization algorithm to determine EGF parameters from a defect-free template image. Defect identification was performed by convolving the sample image with the EGF, achieving 96.30% accuracy on 195 images from the Standard Fabric Defect Glossary dataset. Another approach [8] targeted defect identification in patterned fabrics. Defective blocks were segmented using pattern periodic distance, and compared to a dictionary of features from defect-free blocks. Distance metrics and thresholding were used to identify defects, with accuracy exceeding 95%. An enhanced YOLOv4 architecture [9] incorporated the Soft-Pool layer within the Spatial Pyramid Pooling (SPP) structure. Adaptive histogram equalization was also applied to enhance image quality. Using a dataset that combined Aliyun-FD-10500, Kaggle images, and real photographs, the method achieved a mean accuracy of 86.44%, an improvement of 6% over standard YOLOv4.

A DenseNet-based edge detection algorithm [10] used an optimized cross-entropy loss and six enhancement designs to improve feature representation. The method outperformed conventional CNNs, improving the area under the curve (AUC) by 18% across 11 defect types. A method using direction templates and image pyramids [11] was presented for identifying defects in color and periodically patterned fabrics. A stacked de-noising auto-encoder reconstructed the image from blocks sampled in the pyramid representation. Defective blocks were identified using SSIM, resulting in a mean accuracy of 69.68% on TILDA fabric images. Two CNN-based structures [12] were presented for jacquard-patterned fabrics. One used CNNs for defect identification on isolated patterns, while the other applied integrated state-of-the-art CNNs to the entire dataset. The multispectral dataset included RGB and near-infrared images, which were preprocessed using adaptive histogram equalization.

A saliency-based method [13] identified defects by estimating the membership degree of each defect region. Iterative thresholding and morphological operations were used, achieving a mean identification accuracy of 95.48% on various experimental images. A CNN model [14] was designed to learn defect features from only 50 labeled samples. Initially, the model was applied without training to generate raw outputs, which were later used for supervised learning. Experiments on four fabric datasets with different textures achieved a mean accuracy of

95.89%. A dictionary learning-based method [15] first segmented a defect-free fabric image to build a joint matrix, from which a random dictionary was created. Orthogonal matching pursuit and k-SVD (k singular value decomposition) were applied to learn sparse representations. Defect identification was based on reconstruction error and adaptive thresholding, achieving 93.63% mean accuracy on TILDA classes C_1, C_2, and C_3. A method for nonwoven fabric defect identification [16] was presented using the LL-YOLOv5 network, which incorporated the LSK and Light-RepGFPN modules. These modules enhanced small defect identification and feature fusion. The model achieved a mean accuracy of 90.30% on a hyperspectral nonwoven fabric dataset, outperforming standard YOLOv5 by 2.2%. A Color Conversion Network (CCN) [17] converted RGB images into an optimized color space to better distinguish defects from normal patterns. A contrastive loss function maximized the separation between defect and non-defect features. A complementary adversarial structure, CASDD [18], combined an encoder–decoder module with dual discriminators for identifying texture defects. Edge detection blocks were integrated into convolutional layers, while two discriminators focused on key features and edge differences to improve boundary identification. A YOLO-SCD network [19] used an attention mechanism to enhance feature representation in the neck of the model. It achieved a mean accuracy of 82.92%, improving YOLOv4 performance by 8.49%. Finally, a network incorporating a parallel dilated attention module [20] and a feature pyramid network was introduced to capture multiscale contextual information. Alpha-GIoU loss was used to refine bounding box regression. Additionally, a dual attention module, self-enhanced (SE) and cross-enhanced (SE), was developed [21] to enrich contextual and inter-layer feature representations for improved prediction accuracy.

3- Proposed Method

The steps of the proposed method are illustrated in Fig. 1. Each fabric image is selected from the TILDA dataset and has a resolution of 400×400 pixels. First, the GLCM of the input image is calculated in 0° and 90° orientations. From each matrix, six features are extracted using Equations (1) to (6). The feature defined in Equation (1) captures the scattering of grey levels around the mean intensity.

$$\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(i-\mu)^2\,C(i,j) \tag{1}$$

Where μ, $C(i,j)$, and N_g are respectively mean, GLCM member at (i,j), and a number of grey levels of the image.

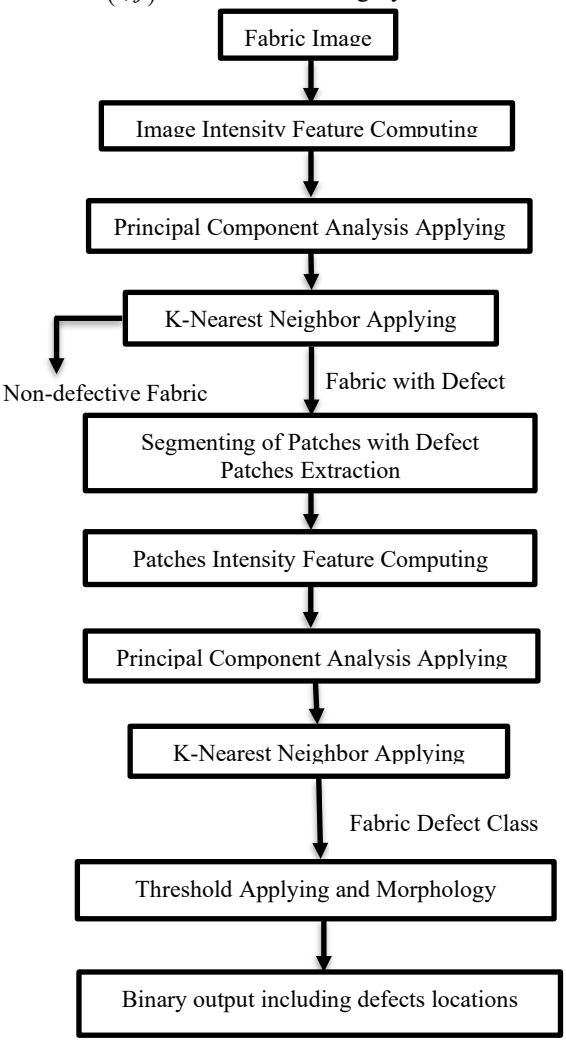

Fig. 1 Flowchart of the proposed method

The feature defined in Equation (2) shows the spreading of sum of grey levels around the mean intensity.

$$C_x(i)=\sum_{j=0}^{N_g-1}C(i,j)$$

$$C_y(i)=\sum_{i=0}^{N_g-1}C(i,j)$$

$$\sum_{i=2}^{2N_g}\left(i-\left[\sum_{i=2}^{2N_g}iC_{x+y}(i)\right]\right)^2 \tag{2}$$

The feature defined in Equation (3) shows the mean distribution of the sum of grey levels.

$$\sum_{i=o}^{2N_g-2}iC_{x+y}(i,j) \tag{3}$$

The feature defined in Equation (4) shows the irregularity of difference distribution of the image intensities.

$$-\sum_{i=0}^{N_g-1} C_{x-y}(i)\log\left(C_{x-y}(i)\right) \quad (4)$$

The feature defined in Equation (5) shows the maximum of the GLCM members.

$$Max_{i,j}\, C(i,j) \quad (5)$$

The feature defined in Equation (6) shows the homogeneity of the grey levels distribution.

$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{C(i,j)}{1+(i-j)^2} \quad (6)$$

In addition to GLCM features, four statistical intensity features, minimum, maximum, median, and mean are computed from the original image.

4- Results and Comparisons

4-1- TILDA Database

TILDA database [27] is a comprehensive database commonly used for evaluating fabric defect identification algorithms. It contains 3,200 images across four classes, including various types of simple and patterned designs. The proposed method was implemented and tested on this dataset. The C_1 class has simple fabrics with a narrow structure. The C_2 class has simple fabrics with a random structure. The C_3 class has periodical structured fabrics. The C_4 class has fabrics with randomly patterns.

4-2- Evaluation Criteria

Four standard evaluation metrics, defined by Equations (7) to (10) [27], are used to assess the performance of the proposed method.

$$Sensitivity\,(Sens.) = \frac{TP}{TP+FN} \quad (7)$$

$$Specificity\,(Spec.) = \frac{TN}{TN+FP} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$False\,Rate\,(FR) = \frac{FP+FN}{TP+TN+FP+FN} \quad (10)$$

Where TP shows true defect identification, which indicates only the pixels with defect are white in the identification result. Both TN and FP show no white pixels in the identification result of the free defect image. FN shows no white pixels in the identification result of the defective image.

4-3- Validation

The simulation of the proposed method is done by *MATLAB* 2019 with 64 bit and operating system of the Windows 7. It is implemented on a system based on the Intel(R) core (TM) i3-2350 CPU @2.3 GHz, 4GB RAM. Several results for C_1 are shown in Table 1. Two groups of designs in C_1 are C_1r_1 and C_1r_3. 200 images of C_1r_1 are defective and 50 images are non-defective. 160 images of C_1r_3 are defective and 50 images are non-defective. The results in Table 1 report the high mean accuracy of the proposed method for defect identification in C_1.

Table 1: Defect identification results of the proposed method for C_1

Class	*Sens.*	*Spec.*	*Accuracy*	*FR*
C_1r_1	90.14%	91%	92.10%	7.90%
C_1r_3	87.94%	100%	91.72%	8.28%

Several results for C_2 are shown in Table 2. Two groups designs in C_2 are C_2r_2 and C_2r_3. 200 images of C_2r_2 are defective and 50 images are non-defective. 200 images of C_2r_3 are defective and 50 images are non-defective. It has to be noticed that black holes are randomly placed in the images background of C_2, that should not be identified as the defects. The results in Table 2 report the high mean accuracy of the proposed method for defect identification in C_2.

Table 2: Defect identification results of the proposed method for C_2

Class	*Sens.*	*Spec.*	*Accuracy*	*FR*
C_2r_2	100%	100%	100%	0%
C_2r_3	97.60%	100%	98.89%	1.11%

Several results for C_3 are shown in Table 3. Two groups designs in C_3 are C_3r_1 and C_3r_3. 170 images of C_3r_1 are defective and 50 images are non-defective. 160 images of C_3r_3 are defective and 50 images are non-defective. The results in Table 3 illustrate that the defect identification in C_3 is more difficult than that on C_2, because of various grey levels in the background of the patterned fabric images in C_3.

Table 3: Defect identification results of the proposed method for C_3

Class	*Sens.*	*Spec.*	*Accuracy*	*FR*
C_3r_1	94.79%	100%	96.53%	3.47%
C_3r_3	93.42%	100%	94.67%	5.33%

4-4- Visual Results

The proposed method several visual results for C_1, C_2, and C_3 are respectively shown in Fig. 2, 3, and 4. In Fig. 2, images in rows 1 to 2 are from C_1r_1, and images in rows 3 to 4 are from C_1r_3. In Fig. 3, images in rows 1 to 2 are from C_2r_2, and images in rows 3 to 4 are from C_2r_3. In Fig. 4, images in rows 1 to 2 are from C_3r_1, and images in rows 3 to 4 are from C_3r_3.

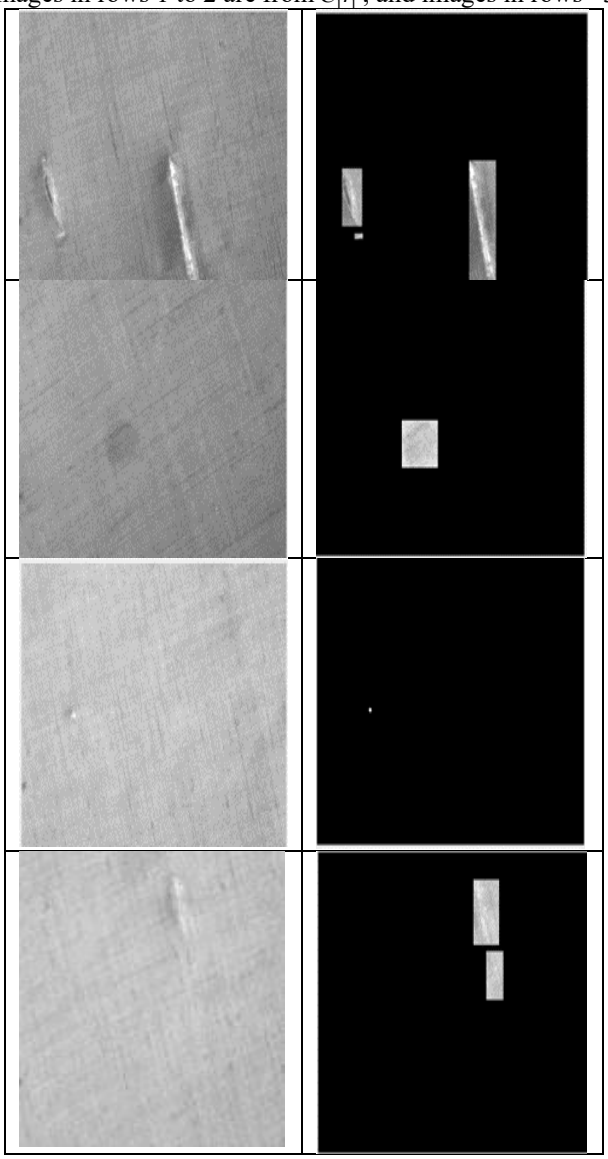

Fig. 2 Several fabric defect identification results for C_1r_1 (rows 1 to 2) and C_1r_3 (rows 3 to 4) by the proposed method

In some images, defects are highly camouflaged within the fabric background, leading to misidentification. Examples of such cases are shown in Fig. 5. Additionally, no results are reported for fabrics with random patterns, as the proposed method performs poorly on this category.

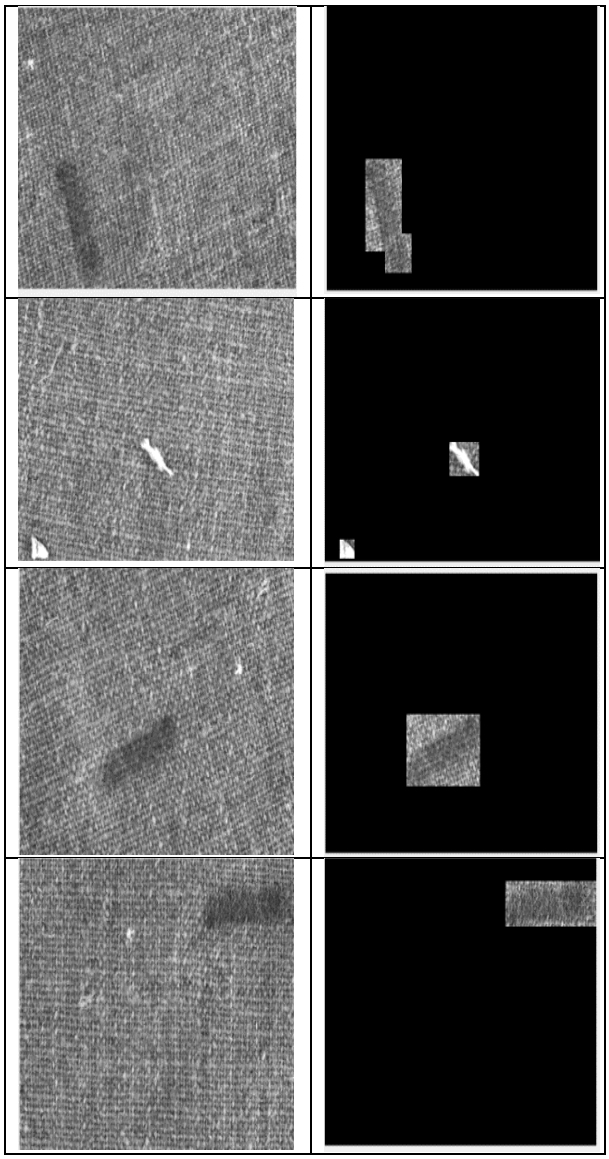

Fig. 3 Several fabric defect identification results for C_2r_2 (rows 1 to 2) and C_2r_3 (rows 3 to 4) by the proposed method

4-5- Comparisons

Several state-of-the-art methods were implemented and evaluated on the TILDA dataset for comparison. Table 4 summarizes the mean defect identification accuracy of each method. While some advanced models achieve slightly higher accuracy, they often rely on deep learning and require complex architectures or large amount of labeled data. In contrast, the proposed method is simple, interpretable, and easy to implement, yet achieves high accuracy for most fabric types, expect randomly patterned fabrics.

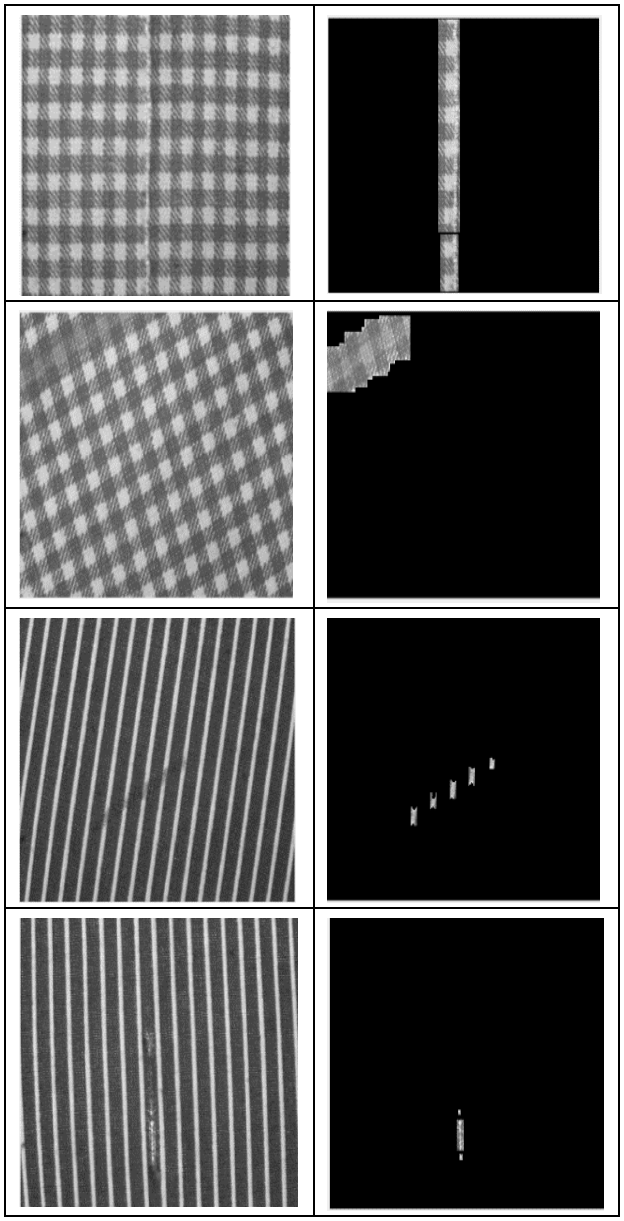

Fig. 4 Several fabric identification results for C_3r_1 (rows 1 to 2) and C_3r_3 (rows 3 to 4**)** by the proposed method

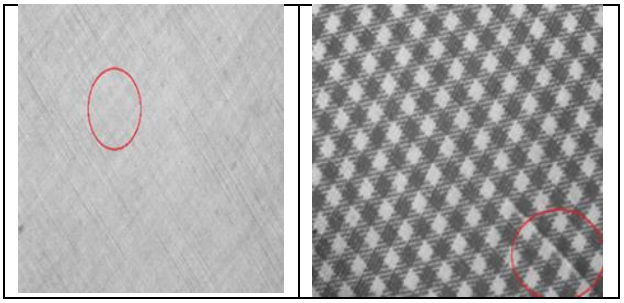

Fig. 5 Two samples of defects fade into the fabric background, leading to misidentification

Table 4: Defect identification mean accuracy of the proposed method compared with a several state-of-the-art methods on the TILDA

Method	Mean Accuracy
[2]	93.70%
[4]	94.95%
[5]	90.62%
[6]	89.86%
[7]	93.84%
[9]	94.37%
[8]	95.75%
[11]	95.25%
[13]	95.68%
[14]	95.30%
[19]	91.97%
Proposed	95.65%

5- Conclusions

This study introduced a simple, supervised method for accurate and reliable fabric defect identification using the KNN classifier and PCA. By reducing the dimensionality of the feature vectors. The proposed method achieves lower memory usage and computational cost, making it suitable for real-time applications.

Extensive experiments on the comprehensive TILDA dataset (3,200 images across four classes) demonstrated the effectiveness of the approach. While the method was not evaluated on fabrics with random patterns, it achieved high identification accuracy on simple and non-random patterned fabrics.

Specifically, the proposed method achieved mean identification accuracies of 91.91% for class C_1, 99.44% for class C_2, and 95.60% for class C_3. The lower performance in class C_1 is attributed to the narrow and less distinguishable textures in simple fabrics. The overall average accuracy across classes C_1 to C_3 is 95.65%.

The key innovation of this work is achieving competitive accuracy without using any deep or non-deep neural network models, relying solely on classical machine learning techniques (PCA and KNN). As a result, the method is well-suited for real-world applications where simplicity, interpretability, and resource efficiency are critical. However, the current method is not applicable to

fabrics with random patterns, as the identification accuracy is insufficient for practical use in such cases.

Given its performance on the TILDA dataset, the method is expected to generalize well to other fabric datasets, provided the fabrics have simple or regular patterns. Future work will focus on extending the method to handle random patterns and more complex fabric structures.

References

[1] S. Mei, Y. Wang, and G. Wen, "Automated Fabric Defect Detection with a Multi-Scale Convolutional De-Noising Auto-Encoder Network Model," Sensors, Vol. 18, No. 4, 2018.

[2] J. F. Jing, H. Ma, and H. H. Zhang, "Automatic fabric defect detection using a deep convolutional neural network," Coloration Technology, Vol. 135, No. 3, 2019, pp. 213–223.

[3] P. Guo, Y. Lin, Y. Wu, R. Hugh Gang, and Y. Li, "Intelligent quality control of surface defects in fabrics: A comprehensive research progress," IEEE Access, Vol. 12, 2024, pp. 63777-63808.

[4] W. Wei, D. Deng, L Zeng, and C. Zhang, "Real-time implementation of fabric defect detection based on variational automatic encoder with structure similarity," Real-Time Image Processing, Vol. 18, No. 3, 2021, pp. 807-823.

[5] L. Shao, E. Zhang, Q. Ma, and M. Lie, "Pixel-wise semi-supervised fabric detection method combined with multitask mean teacher," IEEE Transactions on Instrumentation and Measurement, Vol. 71, 2506011, 2022.

[6] D. Mo, W. K. Wong, Z. Lai, and J. Zhou, "Weighted double-low-rank decomposition with application to fabric defect detection," IEEE Transactions on Automation Science and Engineering, Vol. 18, No.3, 2021, pp. 1170-1190.

[7] Y. Li, H. Luo, M. Yu, G. Jiang, and H. Cong, "Fabric defect detection algorithm using RDPSO-based optimal Gabor filter," The Journal of The Textile Institute, Vol. 110, No. 4, 2019, pp. 487–495.

[8] W. Wang, N. Deng, and B. Xin, "Sequential detection of image defects for patterned fabrics," IEEE Access, Vol. 8, 2020, pp. 174751-174762.

[9] Q. Liu, C. Wang, Y. Li, M. Gao, and J. Li, "A fabric defect detection method based on deep learning," IEEE Access, Vol. 10, 2022, pp. 4284-4296.

[10] Z. Zhu, G. Han, G. Jia, and L. Shu, "Modified DenseNet for automatic fabric defect detection with edge computing for minimizing latency," IEEE Internet of Things Journal, Vol. 7, No. 10, 2020, pp. 9623-9636.

[11] H. Xie, Y. Zhang, and Z. Wu, "Fabric defect detection method combing image pyramid and direction template," IEEE Access, Vol. 7, 2019, pp. 182320-182334.

[12] M. M. Khodier, S. M. Ahmed, and M. Sharaf Sayed, "Complex pattern jacquard fabrics defect detection using convolutional neural networks and multispectral imaging," IEEE Access, Vol. 10, 2022, pp. 10653-10660.

[13] l. Song, r. Li, and S. Chen, "Fabric defect detection based on membership degree of regions," IEEE Access, Vol. 8, 2020, pp. 48752-48760.

[14] Y. Huang, J. Jing, and Z. Wang, "Fabric defect segmentation method based on deep learning," IEEE Transactions on Instrumentation and Measurement, Vol. 70, 2021.

[15] X. Kang, and F. Zhang, "A universal and adaptive fabric defect detection algorithm based on sparse dictionary learning," IEEE Access, Vol. 8, 2020, pp. 221808 - 221830.

[16] H. Lv, H. Zhang, M. Wang, J. Xu, X. Li, and C. Liu, "Hyperspectral imaging based nonwoven fabric-defect-detection method using LL-YOLOv5," IEEE Access, Vol. 12, 2024, pp. 41988-41998.

[17] C. Zhang, Y. Qi, and Y. Wang, " Learning the color space for effective fabric defect detection," IEEE Transactions on Emerging Topics in Computational Intelligence, Vol. 8, No. 1, 2024, pp. 981-991.

[18] S. Tian, P. Huang, H. Ma, J. Wang, X. Zhou, S. Zhang, J. Zhou, R. Huang, and Y. Li, "CASDD: Automatic surface defect detection using a complementary adversarial network," IEEE Sensors Journal, Vol. 22, No. 20, 2024, pp. 19583-19595.

[19] X. Luo, Q. Ni, R. Tao, and Y. Shi, "A lightweight detector based on attention mechanism for fabric defect detection," IEEE Access, Vol. 11, 2023, pp. 33554-33569.

[20] Z. Xiang, Y. Shen, M. Ma, and M. Qian, "HookNet: Efficient multiscale context aggregation for high-accuracy detection of fabric," IEEE Transactions on Instrumentation and Measurement, Vol. 72, 2023.

[21] Y. Zhao, Q. Liu, H. Su, J. Zhang, H. Ma, W. Zou, and S. Liu, "Attention-based multiscale feature fusion for efficient surface defect detection," IEEE Transactions on Instrumentation and Measurement, Vol. 73, 2024.

[22] H. Chugh, M. Garg, S. Gupta, and S. Sharma, "Plant leaf image identification with texture features using microstructure descriptor," in the 10th International Conference on Reliability, Infocom Technologies, and Optimization (Trend and Future Decisions), 2022. DOI:10.1109/ICRITO56286.2022.9965064

[23] Z. Xia, Y. Chen, and C. Xu, "Multiview PCA: a methodology of feature extraction and dimension reduction for high-order data," IEEE Transaction on Cybernetics, Vol. 52, No. 10, 2022, pp. 11068-11080.

[24] F. Nie, Z. Li, R. Wang, and X. Li, "An effective and efficient algorithm for k-means clustering with new formulation," IEEE Transactions on Knowledge and Data Engineering, Vol. 35, No. 4, 2023, pp. 3433-3443.

[25] B. Yang, Z. Jia, J. Yang, and N. K. Kasabov, "Video snow removal based on self-adaptation snow detection and patch-based Gaussian mixture model," IEEE Access, Vol. 8, 2020, pp. 60188-160201.

[26] I. Misra, M. Kumar Rahil, S. Manthira Moorthi, and Debajyoti Dhar, "Direct feature extraction and image co-registration of morphological structure from Bayer pattern raw planetary images," Expert Systems and Application, Vol. 238, 2024.

[27] M. Boluki, and F. Mohanna, "Inspection of textile fabrics based on the optimal Gabor filter," Signal, Image, and Video Processing, Vol. 15, 2021, pp. 1617-1625.

Journal of Information Systems and Telecommunication
Vol.13, No4, October-December 2025, 333-345
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Federated Learning for Privacy-Preserving Intrusion Detection: A Systematic Review, Taxonomy, Challenges and Future Directions

Dattatray Raghunath Kale[1*], Amolkumar N Jadhav[2], Swati Shirke-Deshmukh[3], Sunny Baburao Mohite[2], Shrihari Khatawka[4], Rahul Sonkamble[5], Sarang Patil[6], Madhav Salunkhe[4]

[1].Department of Computer Science & Engineering, MIT Art Design and Technology University, Pune, India
[2].Department of Computer Science & Engineering, D Y Patil College of Engineering and Technology, Kolhapur, India
[3].Department of Computer Science & Engineering, Pimpri Chinchwad University, Pune, Maharashtra, India
[4].Department of Computer Science & Engineering, Annasaheb Dange College of Engineering and Technology, Ashta
[5].Department of Computer Science & Engineering, Pimpri Chinchwad University, Pune, Maharashtra, India
[6].Amity School of Engineering and Technology, Amity University, Mumbai, Maharashtra, India

Received: 08 Feb 2024/ Revised: 04 Dec 2025/ Accepted: 11 Jan 2026

Abstract

This paper presents a systematic review of intrusion detection systems (IDS) that leverage federated learning (FL) to enhance privacy in distributed cybersecurity environments. A total of 78 peer-reviewed studies published between 2019 and 2024 were selected using PRISMA guidelines. We categorize FL-based IDS solutions based on architecture (centralized, decentralized, hierarchical), aggregation methods (e.g., FedAvg, DAFL), and privacy-preserving techniques (e.g., differential privacy, homomorphic encryption). The survey also examines solutions to key challenges such as communication overhead, data heterogeneity, and poisoning attacks. Furthermore, this study outlines unresolved issues and proposes future research directions, including adaptive federated optimization and cross-domain deployments. This review serves as a valuable resource for researchers and practitioners aiming to develop privacy-aware, scalable, and intelligent IDS using federated learning.

Keywords: Federated Learning; Intrusion Detection; Data Privacy; Cyber security.

1- Introduction

Cybersecurity threats continue to grow in complexity and scale, posing significant risks to individuals, organizations, and critical infrastructure. Intrusion Detection Systems (IDS) play a vital role in identifying unauthorized activities and protecting digital assets. While machine learning (ML)-based IDSs have improved detection accuracy, most rely on centralized architectures that require aggregating raw data from multiple sources raising serious privacy concerns. Regulatory frameworks such as GDPR and HIPAA further restrict data sharing across entities. As a result, there is an urgent need for privacy-preserving IDS solutions that can operate effectively across distributed environments without exposing sensitive data.

Cybersecurity threats are a most important problem in today's world, which is becoming more digital and interrelated by the day. They pose a threat not only to individual privacy but also to the working constancy of industries and national infrastructure. Networked system vulnerabilities are often used by malicious actors to get illegal access, bargain confidential data, hinder services, or expose data integrity [1]. These risks encompass a wide variety of attacks, such as ransomware, phishing, denial-of-service (DoS), zero-day exploits, and Advanced Persistent Threats (APTs). Thus, it is more significant than ever to have defense mechanisms that are both smart and active. By continuously seeing system behavior, network traffic, and user actions, intrusion detection systems (IDS) play a vital part in the defense ecosystem by catching infrequent or doubtful patterns that could point to cyber intrusions [2][3]. IDS must progress to become more accurate, flexible, and proactive in present threat detection while reducing false positives and assuring system scalability, seeing the dynamic character of cyberattacks and their growing complexity.

✉ **Dattatray Raghunath Kale**
kaledatta156@gmail.com

Traditional IDS technologies have advanced, especially those that use deep learning (DL) and machine learning (ML) for anomaly detection, but there are still a number of noteworthy matters that necessitate being addressed. The centralized architectures used by the mainstream of ML-based IDS methods combine raw data from several terminations into a single server for training and valuation. However, this pattern presents thoughtful risks to data privacy because it may expose private user data during transmission or storage, including IP logs, user IDs, medical histories, or personal behavioural patterns [4][5]. Administrations are also regularly unable or grudging to share data outside of their locations due to ethical, lawful, and regulatory restrictions like GDPR, HIPAA, and data authority laws. This leads to disjointed datasets with little variety, which makes it harder to detect attacks in real-world surroundings and causes biased learning and poor generalization [6]. Strong intrusion detection model training is additionally difficult due to class inequality, data sparsity, non-IID (non-independent and identically distributed) data distributions, and altering attack signatures. Structuring a safe, supportive, and scalable IDS therefore needs tackling these privacy and data distribution problems.

Federated Learning (FL), which protects user privacy while taking the disadvantages of centralized learning, has become a game-changer. Without sharing raw data, it allows numerous clients like distributed organizations, edge nodes, or IoT devices to work together to train a common global model [7][8]. Private data is kept local and secure because only model updates such as weights or inclines are sent. IDS applications, where privacy and security are vital, are preferably right for this distributed learning framework. FL is extremely applicable to businesses like finance, healthcare, perilous infrastructure, and smart cities because it permits administrations to gain from shared knowledge and model development without exposing sensitive data [9][10][11]. Additionally, new progress in FL includes privacy-enhancing skills such as secure multiparty computation (SMC), homomorphic encryption, blockchain-based authentication, and differential privacy [12][13]. These protections raise confidence, lower the possibility of privacy destruction, and promise robust protection against aggressive movements like model inversion and data poisoning. FL thus encourages a cooperative cybersecurity ecosystem in addition to addressing the disadvantages of data sharing [14][15].

Even though there is a rising amount of study on the use of FL in IDS, there are still a number of noteworthy gaps. A detailed framework for comparing FL-IDS models across significant sizes, including model aggregation strategies (e.g., FedAvg, FedProx, DAFL), privacy-preserving techniques (e.g., differential privacy, SMC), system architectures (e.g., centralized vs. hierarchical FL), and practical deployment circumstances, is missing in many formerly published works that focus on developing particular FL algorithms or privacy techniques [16][17]. Moreover, the mainstream of study disregards the trade-offs between accuracy, latency, communication overhead, and resource consumption, all of which are dangerous for applying FL in surroundings with partial resources, like edge networks and the Internet of Things [18][19]. It is also challenging for experts and investigators to accept or enlarge upon current solutions due to the lack of discussion surrounding benchmark datasets, performance evaluation metrics, and scalability across various domains. An embattled, inclusive, and prepared investigation of FL precisely within the IDS domain is lacking, despite the fact that previous reviews have examined FL and IDS autonomously [20]. By presenting a thorough literature review that highlights problems, classifies approaches, and proposes future research paths, this work fills that knowledge gap.

FL is a machine learning method that enables the development of models across decentralized edge computers or servers holding local data samples without requiring data exchange. This paradigm for collaborative learning is especially pertinent when discussing privacy issues with intrusion detection systems (IDS). For efficient training and pattern recognition, traditional intrusion detection systems frequently need access to private and sensitive data. However, there are serious privacy concerns associated with gathering and centralizing such data. FL offers a promising solution by allowing ML models to be trained on distributed data without requiring central data sharing.

This paper's major goal is to offer a wide and systematic overview of the state-of-the-art in Federated Learning for Intrusion Detection Systems (FL-IDS) from 2019 to 2024. Using prearranged presence and elimination criteria, 78 peer-reviewed articles in all were selected from reputable digital libraries, including IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink.

The following are the contributions made by this survey:

- System architecture, combination plans, privacy mechanisms, and application areas (such as IoT, IIoT, 5G, and healthcare) are used to classify FL-IDS methods.
- It examines methods for refining privacy, such as adversarial defense, differential privacy, and secure aggregation.
- It highlights significant problems and restrictions like federated poisoning attacks, data heterogeneity, and communication overhead.
- With an emphasis on edge computing, cross-domain transfer, adaptive FL models, and real-time IDS deployment, it gives a research roadmap and future directions.

This paper's residual segments are arranged as follows: The paper is resolved with final perceptions and practical implications in Section 6. Section 4 deliberates significant open challenges; Section 5 summarizes future research directions; Section 3 analyses and classifies existing FL-IDS methods; and Section 2 presents the basic ideas of IDS and FL and highlights privacy challenges.

2- Foundations and Literature Overview

FL entails cooperatively updating models across decentralized devices. Let's represent the key components, Global Model parameters θ and Local model parameters for device i is θi The global objective function F(θ) is typically defined as the average of local objective functions across all devices:

$$F(\theta) = \frac{1}{n}\sum_{i=0}^{n} f(\theta i) \quad (1)$$

Here n is the overall quantity of devices, $f(\theta i)$ shows local objective function for device $\boldsymbol{i}$ At each iteration, each device $\boldsymbol{i}$ computes a local update Δθi by minimizing its local objective function as

$$\Delta\theta_i = arg\ min\ \Delta\theta\ fi(\theta_i + \Delta\theta_i) \quad (2)$$

The local model updates $\Delta\theta_i$ subsequently communicated to a centralized server for compilation and global models updated by aggregating the local updates as,

$$\theta \leftarrow \frac{1}{n}\sum_{i=1}^{n}(\theta_i + \Delta\,\theta_i) \quad (3)$$

The process of local updates, communication, and aggregation is repeated for multiple iterations or until convergence.

Federated Learning (FL), which redefines the conventional data-centric techniques, emerges as a promising paradigm that addresses these issues and changes the intrusion detection landscape [21]. In federated learning, businesses, manufacturers, or devices that interact with data are considered clients, and each client's data privacy is preserved [22]. Clients build identical deep local models and train them with their own datasets. On the cloud center server, create a global depth model with the same framework as the local model [23]. Through constant communication between the training's central server and several clients, the global and local models are transferred as shown in figure1. In order to accomplish particular learning tasks, a global depth model with outstanding performance is ultimately jointly established. The two primary stages of a federated learning scenario are local update and global aggregation, to put it briefly [24]. This illustrates how clients can share and profit from each other's data through FL without having to send private information to a central server. Federated Learning, with its decentralized model training paradigm, provides a novel solution that prioritizes protecting sensitive user data privacy while simultaneously improving intrusion detection models' accuracy and efficiency. In this extensive analysis, we examine the complementary nature of FL and IDS, delving into the subtleties of this innovative technology and its revolutionary effects on cybersecurity privacy protection [25][26].

The purpose of this survey is to present an in-depth study of the difficulties posed by centralized intrusion detection systems (IDS) models, highlighting the privacy implications that are now a major topic in the discussion of network security [27][28]. By keeping aim to clarify how this decentralized learning paradigm minimizes privacy concerns while preserving intrusion detection effectiveness by delving into the core ideas of federated learning [29][30]. This paper a look into the future where the combination of decentralized machine learning and cybersecurity strengthens digital defences and upholds the fundamental right to privacy in an increasingly connected world as we navigate the complexities of Federated Learning for enhanced privacy in Intrusion Detection Systems [31]. This is a journey that goes beyond traditional boundaries.

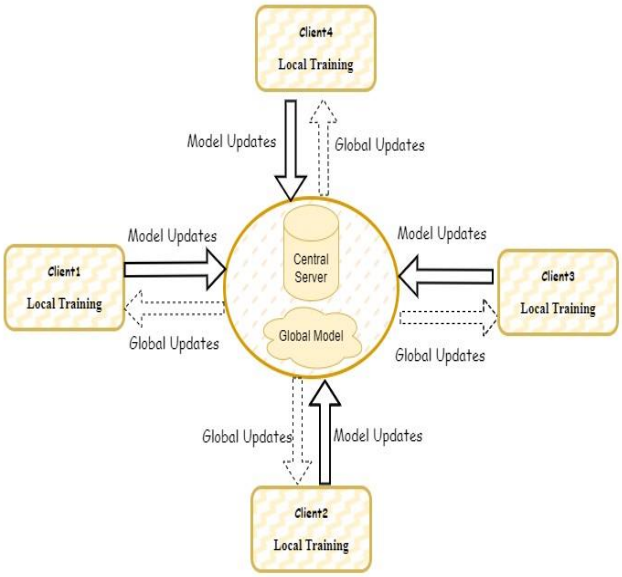

Fig1: Federated Learning System in Generalized Format

Table-1 provides a summary of the definitions of the abbreviations used in the paper in order to aid with comprehension.

Table 1: Common abbreviations listed with explanations

Acronym	*Definition*
FL	Federated Learning
IDS	Intrusion Detection Systems
NIDS	Network Intrusion Detection systems

DAFL	Dynamic Weighted Aggregation Federated Learning
FPR	False Positive Rates
TPR	True Positive Rates
IOT	Internet of Things
DAFL	Dynamic Weighted Aggregation Federated Learning
SMC	Secure Multiparty Computation
ACGAN	Auxiliary Classifier Generative Adversarial Networks

2-1- Federated Learning (FL) Applications in Cyber Security

FL has garnered significant interest in the field of cybersecurity, particularly in relation to intrusion detection systems (IDS). FL allows collaborative learning while preserving data confidentiality and locality. In order to prevent data privacy violations, a novel architecture known as Decentralized and Online Federated Learning Intrusion Detection (DOF-ID) has been proposed. This architecture enables each intrusion detection system to learn from expertise obtained in other systems [32]. An alternative method is the DAFL scheme, which better detects intrusions with less communication overhead by implementing adaptive selection and balancing strategies for local models. [33]. In this regard, FL has great promise, as demonstrated by Campos [34] and Ferrag [35], who particularly point out that FL is more accurate and private than non-federated learning. Alazab [36] highlights the technology's potential for real-time cybersecurity by offering a thorough overview of FL models for authentication, privacy, trust management, and attack detection. A hybrid ensemble approach for FL-based IDS in IoT security is presented by Chatterjee [37], which achieves low FPR and high TPR on both clean and noisy data.

2-2- Privacy-Enhancing Techniques in FL-IDS

According to Ruzafa-Alcazar's [38] evaluation of differential privacy techniques, using Fed+ yields result that are comparable to those of non-privacy-preserving techniques. But it does not provide a comprehensive analysis of the communication and computational overhead associated with the application of such techniques in resource-constrained IIoT scenarios. When Ansam Khraisat [39] evaluates Federated Learning against conventional deep learning models, Federated Learning outperforms the latter in terms of accuracy and loss, especially in situations where data security and privacy are prioritized. Federated mimic learning, a novel approach put forth by Al-Marri [40], mixes mimic learning and federated learning to produce a distributed intrusion detection system that poses the least risk to users' privacy but this research has several shortcomings: it does not examine potential vulnerabilities, does not compare the suggested solution with other privacy-preserving methods, and raises scalability issues. It also does not address computational and communication overhead. In 2020, Yang presents privacy-preserving protocols that use cryptographic techniques to safeguard participant parameter data in Federated Learning [41]. To safeguard identity privacy, a lightweight linkable ring signature scheme is suggested in [42].

Among the multiple techniques, one technique is to securely compute patient-level similarity scores amongst hospitals using Secure Multiparty Computation (SMC), which allows patient clustering without sharing patient-level data [43]. In order to ensure privacy guarantees, the Federated Learning framework incorporates differential privacy (DP), which involves adding calibrated noises. This approach has been applied to the Federated Averaging algorithm, resulting in the ULQ-DP-FedAvg [44]. Additionally, the Fed+ aggregation function produced comparable results even with the addition of noise to the federated training process when differential privacy techniques were evaluated for training a FL-enabled IDS for industrial IoT [45]. These methods seek to protect sensitive data in Federated Learning for IDS while solving issues related to privacy.

The effectiveness of FL in IDS has been evaluated taking into account data heterogeneity, non-independent and identically distributed (non-IID) data, and data privacy concerns. One study found that non-IID data had an impact on FL performance and proposed a FL data rebalancing method based on ACGAN [46]. An additional study evaluated the effectiveness of FL IDS solutions with respect to deep neural networks (DNNs) and deep belief networks (DBNs) using a realistic dataset of IoT network traffic. In order to lessen the effects of data heterogeneity, they investigated pre-training and different aggregation techniques [47]. An MCDM framework was also developed in order to standardize and benchmark ML-based IDSs utilized in FL structure for IoT app development. The framework included standardizing assessment criteria, developing an evaluation decision matrix, benchmarking, and using MCDM techniques to select the best IDSs. [48].

The application of Hierarchical Federated Learning (HFL) and Federated Averaging (FedAvg) to enhance the speed and precision of Intrusion Detection Systems (IDS) in Internet of Things applications is highlighted by Saadat [49] and Lazzarini [50]. However, Federated Adaptive Gradient Methods (Federated AGMs) are presented by Tong [51] as a possible advancement over current techniques, especially when handling non-IID and unbalanced data. More emphasis is placed on the necessity of ongoing study and analysis of various approaches in practical settings by Campos [52], particularly in the

framework of IoT. Furthermore, distinct data rebalancing strategies and aggregation techniques, like auxiliary classifier generative adversarial networks (ACGAN), can lessen the detrimental effects of non-IID data on FL. To deal with the heterogeneity of data, models, and computation, Full Heterogeneous Federated Learning (FHFL) creates synthetic data, aggregates models using knowledge distillation, and makes use of idle computing resources [53]. These approaches allow for cooperative model training in FL for IDS while maintaining privacy protection. Table-2 lists the different datasets that have been used in previous research.

Table-2 The data sets used in related research

Datasets
NSL-KDD dataset: 125,973 training records, 22,544 test records, used for network intrusion detection
ToN_IoT dataset: 83 features, 4,404,084 samples
NSL-KDD dataset
NSL-KDD dataset for Federated Learning in IoT intrusion detection evaluation.
MNIST and UCI Human Activity Recognition Dataset
Bot-IoT dataset, the MQTTset dataset, and the TON_IoT dataset

In terms of reliability, effectiveness, and adaptability, federated learning (FL) has shown promising results for intrusion detection system (IDS) applications. Without distributing the raw data, FL allows for ML/DL models to be trained on network traffic data gathered from several edge devices [54]. Internet of Medical Things (IoMTs) and tactical military environments simply a few of the domains where FL has been successfully applied [55][56]. With accuracy rates exceeding 93%, FL has demonstrated high model performance in identifying malicious activity. By using federated training of local device data, FL further guarantees data security and privacy while maintaining privacy and improving the model as a whole. Further demonstrating FL's efficiency is the fact that it achieves good detection performance with little network communication overhead. Marulli [57] underscored the significance of efficiency and effectiveness in Federated Learning (FL). Specifically, she stressed the need for accurate federated algorithm tuning and evaluated the trade-offs between accuracy decay and latency in a decentralized learning approach. These results demonstrate FL's potential as a useful strategy for intrusion detection systems (IDS) applications, providing precise identification, effective communication, and scalability in a variety of network environments. All of these studies highlight FL's potential in IDS applications, but they also point to the need for more research to maximize FL's effectiveness.

2-3- Compare FL-IDS with traditional centralized IDS models

This paper compares traditional centralized IDS models with FL-IDS, a decentralized framework for federated learning (FL) with authentication and verification. FL-IDS uses blockchain technology to manage identities dynamically and stops unauthorized parties from initiating poisoning attacks [58]. It permits local devices to confirm the received global model and guarantees that only authorized local devices can add updates to the blockchain. Traditional centralized IDS models, on the other hand, are vulnerable to single points of failure because they depend on centralized servers. FL-IDS provides decentralization, non-tampering, and non-counterfeiting benefits by substituting blockchain technology for the centralized server in order to address this problem [59]. Furthermore, compared to conventional algorithms, FL-IDS is demonstrated to be more communication-efficient and resilient against malevolent nodes [60]. Together, these studies highlight FL's potential to improve IDS privacy and performance in cybersecurity. Table 3 shows the comparison of FL vs Centralized IDS.

Table-3 Comparison Table: FL vs Centralized IDS

Feature	***Traditional Centralized IDS***	***FL-based IDS***
Data Sharing	Requires sending raw data to server	Only model updates shared
Privacy	Low (data exposure risk)	High (local data stays private)
Scalability	Moderate	High (edge-device friendly)
Resilience	Vulnerable to single point failure	Decentralized and more robust
Communication Cost	Low (single server)	High (needs efficient compression)
Security	Central server is target	Can include secure aggregation

Table 4 summarizes key federated learning approaches applied in IDS research between 2020 and 2024, highlighting their contributions and limitations.

Table-4 Comparative Summary of Federated Learning-Based IDS Approaches

Ref	***Year***	***Methodology & Advantages***	***Drawbacks***
[33]	2023	DAFL - Adaptive model selection to reduce communication	Needs optimization for real-time IoT scenarios
[36]	2022	Overview of FL in IDS for privacy & trust management	Lacks model-specific evaluation
[38]	2023	Fed+ using DP for IIoT with comparable results to non-FL	Overhead not discussed for low-resource settings

[40]	2020	Federated mimic learning for distributed IDS	Limited scalability, lacks comparison with other techniques
[46]	2023	Data rebalancing using ACGAN to handle non-IID data	Increased model complexity

3- Methodology

3-1- Search period and rationale for the 2024 cutoff

The review covers studies published from January 2019 up to March 2024. The cutoff date (March 2024) reflects the date when the systematic search and data extraction pipeline were executed. This ensures consistency and reproducibility. We explicitly note that newer works published after March 2024 are not included but may be incorporated in a future update.

3-2- Databases and justification

We selected IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, and arXiv as the primary sources. These were chosen due to their wide coverage of peer-reviewed ML and cybersecurity research and inclusion of both published and preprint works.

3-3- PRISMA Diagram

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) rules helped as the motivation for this survey's systematic method, which guarantees thorough exposure, reproducibility, and transparency. The review focused on the study that addressed Federated Learning (FL) in the context of Intrusion Detection Systems (IDS) and was published between January 2019 and March 2024. It paid specific attention to model presentation and privacy-preserving procedures. A planned search was carried out across five main academic databases, like ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect, and arXiv, to collect relevant works. The search terms (e.g., "Federated Learning" OR "FL") AND ("Intrusion Detection System" OR "IDS") AND ("privacy" OR "cybersecurity" OR "non-IID" OR "aggregation") are collective keywords and Boolean operators. The 148 papers that were reimbursed by the original search were riddled and divided into three steps: (1) full-text review, (2) abstract showing, and (3) duplicate removal. Following the application of the exclusion criteria (non-English, editorial/commentary papers, or general ML unrelated to IDS) and inclusion criteria (peer-reviewed, focused on FL-IDS, practical relevance, and investigational detail), 78 studies in total were selected for additional investigation. The identification, showing, suitability, and insertion phases of the selection process were defined in a PRISMA flow diagram. A PRISMA flow diagram illustrating the review process has been included in the revised version as Figure 2.

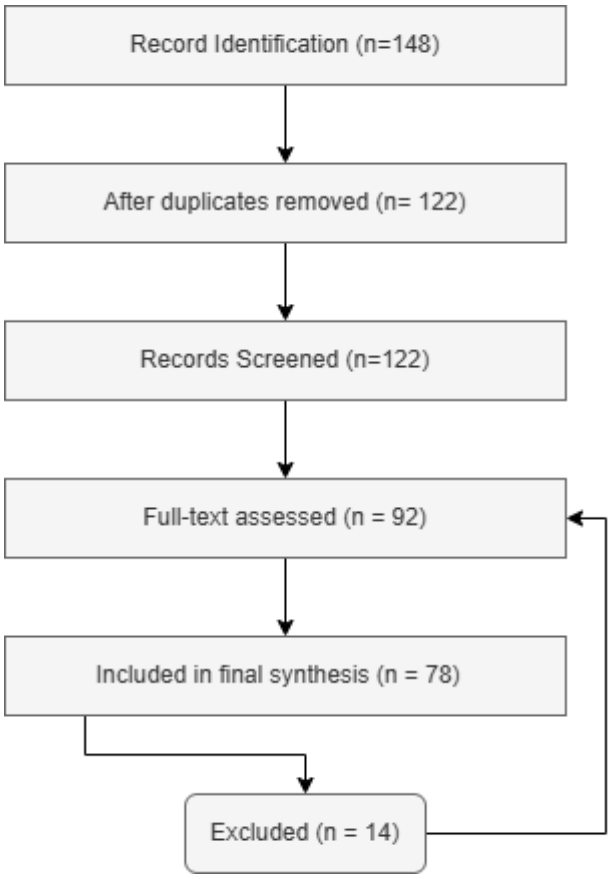

Fig2. PRISMA Flow Diagram

3-4- Screening and CASP checklist and Parisa code explanation

Screening was performed by two independent reviewers, with discrepancies resolved by a third adjudicator. Both reviewers applied the CASP checklist to assess study quality. 78 studies met inclusion criteria, each satisfying at least 5 of the 7 CASP key items. A coding context covering publication details, FL architecture (centralized, hierarchical, decentralized), privacy-enhancing methods (e.g., secure multiparty computation, blockchain integration, differential privacy), aggregation strategies (e.g., FedAvg, DAFL, FedCME), datasets used (e.g., NSL-KDD, ToN-IoT, CSE-CIC-IDS2018), and performance metrics (accuracy, precision, recall, FPR, TPR) was used to thoroughly extract data from the selected revisions. To find tendencies, technical progressions, practical uses, and

research gaps, the studies were assembled and studied thematically. An improved form of the CASP (Critical Appraisal Skills Programme) list was used to measure the objective clarity, experimental consistency, significance of results, transparency of procedure, and discussion of limitations to regulate the reliability and procedural consistency of the included works. Articles were only involved if they pleased a minimum quality standard in each of these extents. This systematic and detailed method guarantees that the survey delivers a reliable and insightful summary of the varying arena of FL-based intrusion detection systems. The CASP checklist template and scoring thresholds used in this study are provided in Appendix B.

Parisa v1.0 is a Python-based automation script used to standardize the search and extraction process across databases. It uses libraries such as requests, pandas, and pyPDF2 to automate query execution and deduplication. All inclusion/exclusion decisions were made by human reviewers. The Parisa repository can be shared upon request.

4- Findings

4-1- Communication Overhead

The transmission of model parameters in every round of FL-based methods results in high communication costs that can impede their actual deployment and pose security risks [61]. Additionally, FL training is negatively impacted by the large model size and equally dispersed private data, particularly in distillation-based FL [62]. In order to tackle these difficulties, scholars have suggested techniques like semisupervised FL through knowledge distillation and DAFL. To enhance detection performance and minimize communication overhead, these techniques make use of unlabelled data, adaptive filtering and balancing strategies for local models, and optimized deep neural networks [63]. Based on experimental results, these methods are effective in improving detection performance while requiring less communication overhead.

4-2- Heterogeneity

Heterogeneity presents challenges for federated learning (FL) in intrusion detection systems. The heterogeneity of data in FL can lead to slower convergence speed, affecting model performance [64]. The training of FL models may also be hampered by non-IID data, which is frequently found in IoT systems [65]. Many methods have been suggested to deal with these issues. One method is to train local models with non-IID data using instance-based transfer learning [66]. An alternative strategy for reducing the effects of data heterogeneity is to make use of pre-training and investigate various aggregation techniques [67].

4-3- Federated Poisoning Attacks

Federated poisoning attacks pose a challenge in FL for IDS. Federated architectures work better because of the distributed nature of data found in client edge devices. Although this property protects the privacy of the data while it's in transit and keeps it from being gathered in one location, the data in question is still at risk. The labels of the data can be readily changed on a client's device. We refer to these attacks as poisoning attacks. These attacks compromise the global model's accuracy and privacy by having malevolent actors alter training data or model updates. In order to address this issue, several papers suggest defence mechanisms against poisoning attacks. Wang et al. propose a PAPI-attack that exploits distinctive capacity in cyclical model updates to infer sensitive information [68]. Yan et al. introduce a CLP-aware defence against poisoning of federated learning (DeFL) that detects malicious clients and identifies critical learning periods to guide the removal of detected attackers [69]. To stop data poisoning attacks, Ovi et al. provide a confident federated learning framework that verifies label quality and removes incorrectly labelled samples from local training [70].

Addressing these challenges requires a combination of algorithmic advancements, technological solutions, and robust privacy-preserving mechanisms. Ongoing research and development efforts are focused on overcoming these obstacles and improving the practicality of Federated Learning in various applications.

4-4- Future Directions

Federated learning is a vibrant and developing field of study. There are still many important new areas that need to be investigated, even though recent work has started to address the issues covered in Section of challenges. We briefly discuss a few promising research directions in the context of privacy-centric intrusion detection systems. Future directions entail investigating and addressing emerging challenges, integrating cutting-edge technologies, and improving the useful applicability of federated learning. Future directions that could be pursued are as follows:

Effective Model Transfer and Compression:

In order to minimize communication overhead in federated learning, investigate methods for effective model compression and transfer. Given the complexity and heterogeneity of network traffic generated by distributed networks such as wearables, mobile phones, and

autonomous vehicles, privacy-preserving decentralized learning techniques like federated learning (FL) have become essential. In order to train a model collaboratively across multiple institutions without requiring local data sharing, FL ensures both privacy and security. Unfortunately, domain feature shift brought on by various acquisition devices or clients can impair the performance of FL models. In response, a brand-new trusted federated disentangling network known as TrFedDis has been put forth. It makes use of feature disentangling to preserve local client-specific feature learning and capture global domain-invariant cross-client representation on the one hand. [71].

Flexible and Adaptive Federated Learning:

Create flexible and dynamic federated learning frameworks that can adapt to evolving intrusion patterns and network conditions. This could entail developing self-learning models that can adjust on their own to changing network topologies and novel forms of attacks. Large training iterations, a lack of adaptivity, and non-IID data distribution are just a few of the difficulties encountered in federated learning that have been brought to light by existing research in this field. Several papers propose adaptive algorithms that address these challenges and provide theoretical guarantees for convergence and improved performance. For example, Kim et al. propose Δ-SGD a step size rule for stochastic gradient descent (SGD) that enables each client to use its own step size based on the local smoothness of the function being optimized [72]. Furthermore, a dynamic adaptive cluster federated learning scheme is put forth to handle changes in real-time data distribution and offer flexibility in cluster partitioning [73]. These approaches demonstrate the importance of flexibility and adaptivity in FL-IDS.

FL's Encryption Standards:

In order to further improve the protection of sensitive data during the federated learning process, research and put into practice advanced privacy-preserving mechanisms like homomorphic encryption, which encrypts local gradients or model updates before they are shared with the centralized server [74], safe multi-party calculations [75], and separate privacy.

Cross-Domain Federated Learning:

Extend research into cross-domain federated learning, where models trained in one domain can be applied to enhance intrusion detection in a different domain. This methodology has been implemented across multiple fields, such as 2D surgical image segmentation [76] and knowledge graph embedding [77]. Regarding surgical image segmentation, the technique tackles issues of data scarcity, privacy safeguarding, and domain shifts between various canters. The method improves the embedding of various clients in knowledge graph embedding by facilitating safe interaction between domains without requiring data sharing.

Edge Computing in FL based IDS:

Federated learning is incorporating edge computing for intrusion detection systems (IDS). This method shifts model aggregation to edge servers in order to preserve data privacy and enhance federated learning performance. [78]. In C-V2X networks, edge computing has greatly improved Intrusion Detection System (IDS) performance, especially when paired with Federated Learning [79]. Resource-efficient FL techniques, such as knowledge distillation and model compression, have been investigated within the framework of mobile edge computing in order to meet the demanding resource requirements of mobile clients [80].

By exploring these research methods, the field of FL-IDS can progress toward intrusion detection systems that are more resilient, flexible, and privacy-preserving, and that are better suited to handle the changing demands of cybersecurity.

5- Results and Synthesis

This section précises the main conclusions drawn from a detailed investigation of 70 chosen papers on Federated Learning (FL) for Intrusion Detection Systems (IDS). The consequences are prepared into two main themes: (1) the state of FL-IDS architectures and privacy policies at the moment, (2) remarkable model contributions like TrFedDis.

5-1- Summary of Trends in FL-Based IDS Research

According to the analysis, decentralized and privacy-preserving IDS designs driven by FL are becoming more and more common, particularly in the IoT, IIoT, and healthcare environments. FedAvg is still the most popular aggregation method, with FedProx, DAFL, and FedCME succeeding thoroughly behind. Each of these methods handles a diverse set of matters, such as communication bottlenecks and client heterogeneity. Differential privacy, secure multiparty computation (SMC), and blockchain integration are examples of privacy-enhancing methods that have increased in popularity because they offer layered defense in contradiction of adversarial attacks and data leakage. Studies are beginning to highlight the trade-offs between resource consumption, system latency, and detection accuracy.

5-2- Performance and Contribution of the TrFedDis Model

Among the latest growths, the Trusted Federated Disentangling Network (TrFedDis) stands out as a prominent model that handles the problem of non-IID data distributions and domain feature change. TrFedDis uses feature separation to maintain local-specific illustrations while learning domain-invariant features across clients, in contrast to traditional FL models that experience performance deprivation as a result of client heterogeneity. According to experimental assessments, TrFedDis outperforms standard FedAvg and FedProx in non-IID environments by up to 6% in terms of accuracy. Additionally, by assuring confidence-aware aggregation, it progresses flexibility against poisoning attacks. By enhancing generalizability and trust in global model updates, features crucial for practical arrangements in dynamic surroundings, this model makes a considerable contribution to the FL-IDS domain.

5-3- Revisited Concepts with Deeper Insights

Our study shows delicate transformations in the applicability of methods like adaptive clustering, knowledge distillation, and model compression, which are usually deliberated across studies. Model compression approaches like quantization and thinning work best in surroundings with inadequate resources, such as mobile edge devices. When models are moved across heterogeneous devices or between domains, knowledge distillation helps to preserve performance. For managing non-IID data and enhancing fairness in cooperative training, adaptive learning methods such as clustered FL and personalized FL present feasible responses. However, the effectiveness of these approaches is regularly determined by the particular IDS application domain and infrastructure limitations.

6- Conclusions

This review thoroughly investigates the use of Federated Learning (FL) in Intrusion Detection Systems (IDS), providing an organized taxonomy and deep analysis of key challenges and solutions. By addressing privacy concerns, communication constraints, and data heterogeneity, FL presents a scalable and privacy-aware approach for real-world IDS deployments. The paper also highlights future research directions such as cross-domain FL, adaptive clustering, model compression, and privacy-enhancing encryption standards. These insights offer valuable guidance for researchers and developers working on privacy-centric, distributed intrusion detection solutions across critical sectors like healthcare, smart grids, and IoT-enabled environments.

Appendix

Appendix A: Correction Table

Reviewer Comment	Author Modification
Clarify the reason for selecting articles only until 2024, although we are now at the end of 2025.	Added a paragraph in Section 3.1 – Search period and rationale for 2024 cutoff explaining that the search and extraction were completed in March 2024, hence studies up to that date were included. Also mentioned that future updates will incorporate post-2024 studies.
Present the introduction under a single heading, without internal subdivisions.	Reorganized the Introduction into a single unified section, merging previously separate subsections (1.1–1.3). All uncited statements now include references. (Page 2)
Add references for all uncited statements in the introduction.	Inserted supporting citations for every factual statement about IDS, FL, and privacy challenges. (Pages 2–3)
Move the correction table to the appendices.	Moved the correction table to Appendix A at the end of the paper and added a reference in the text indicating its new location.
In the methodology, explain (with references) the use of the Parisa version code and justify the choice of databases.	Added a new subsection titled “Parisa code explanation” in Methodology (3.4) describing its purpose, implementation (Python 3.8 with requests, pandas, pyPDF2), and manual verification process. Also justified database selection (IEEE Xplore, ACM, SpringerLink, ScienceDirect, arXiv).
Provide a PRISMA diagram.	Included an PRISMA diagram (Figure 2) summarizing identification, screening, and inclusion steps. Added note that a high-resolution image will appear in the camera-ready version.
Indicate who	Added a detailed paragraph in

completed the CASP checklist and how many articles met the inclusion criteria.	Section 3.3 – Screening and CASP checklist specifying that two independent reviewers completed the CASP checklist; 78 studies met inclusion criteria after resolving 8 disagreements. (Page 6)
Include challenges, open issues, and future directions as subsections of the findings.	Restructured Findings (Section 4) to include three new subsections each supported with citations. (Pages 9–11)
Ensure that findings correspond to the selected articles and include proper citations.	Revised Findings and Results & Synthesis sections to ensure every statement is backed by the 78 reviewed studies, with updated in-text references. (Pages 9–12)

Appendix B: CASP Checklist Template and Scoring Thresholds

CASP Question	Response (Yes/No/Partial)	Score
Did the study address a clearly focused issue?	Yes	1
Was the cohort recruited in an acceptable way?	Yes	1
Was the exposure accurately measured?	Partial	0.5
Were the confounding factors identified and accounted for?	Yes	1
Was the follow-up complete and long enough?	Yes	1
Were the outcomes measured in a valid and reliable way?	Yes	1
Overall, was the study of high quality?	Yes	1

Scoring Thresholds: High Quality: 8–10 Moderate Quality: 5–7, Low Quality: 0–4

Appendix C: Parisa code summary and access details

Section	**Description**	**Details**
Code Name	Parisa	Python-based framework for federated intrusion detection system (IDS)
Purpose	Implements privacy-preserving intrusion detection using federated learning	Designed to detect network anomalies across distributed nodes without sharing raw data
Main Features	Federated model training across multiple clients - Local data preprocessing and feature extraction - Model aggregation at central server - Explainable intrusion alerts and risk scores	Supports common network datasets (NSL-KDD, CICIDS2017)
Dependencies	TensorFlow >= 2.12 - PySyft >= 0.7 - pandas >= 2.1 - scikit-learn >= 1.2	Installable via pip
Usage Summary	1. Configure federated clients and server 2. Load and preprocess network datasets locally 3. Train local models and perform federated aggregation	Sample scripts and configuration templates provided in GitHub repository
Access	GitHub Repository: https://github.com/YourUsername/Parisa-FL-IDS	Public access

Acknowledgments

Insert acknowledgment, if any. The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank ... .”

Instead, write “F. A. Author thanks ... .” Sponsor and financial support acknowledgments are also placed here.

References

[1] K. Kurniabudi, B. Purnama, S. Sharipuddin, D. Darmawijoyo, D. Stiawan, S. Samsuryadi, A. Heryanto, and R. Budiarto, “Network anomaly detection research: A survey,” Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 7, no. 2, pp. 1–10, 2019.

[2] I. Manan, F. Rehman, H. Sharif, C. N. Ali, R. R. Ali, and A. Liaqat, “Cyber security intrusion detection using deep learning approaches and Bot-IoT dataset,” in *Proc. 2023 4th Int. Conf. on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, 2023, pp. **1–5**

[3] J. Lánský, S. Ali, M. Mohammadi, M. K. Majeed, S. H. Karim, S. Rashidi, M. Hosseinzadeh, and A. M. Rahmani, “Deep learning-based intrusion detection systems: A systematic review,” *IEEE Access*, vol. **9**, pp. **101574–101599**, 2021.

[4] S. Tyagi, I. S. Rajput, and R. Pandey, “Federated learning: Applications, security hazards and defense measures,” in *Proc. 2023 Int. Conf. on Device Intelligence, Computing and Communication Technologies (DICCT)*, 2023, pp. **477–482**

[5] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” arXiv preprint arXiv:1610.05492, 2016.

[6] T. Li, A. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” *IEEE Signal Processing Magazine*, vol. **37**, no. **3**, pp. **50–60**, May 2019.

[7] D. A. Kumar and S. R. Venugopalan, “Intrusion detection systems: A review,” *Int. J. Adv. Res. Comput. Sci.*, vol. **8**, no. **5**, pp. **356–370**, 201

[8] K. B. Gan, “Intrusion detection systems: Principles and perspectives,” *J. Multidisciplinary Eng. Sci. Studies (JMESS)*, vol. **4**, no. **11**, pp. —, Nov. 2018.

[9] T. F. Lunt, “Foundations for intrusion detection,” in *Proc. IEEE Computer Security Foundations Workshop (CSFW)*, 2000, pp

[10] S. Mukkamala, A. H. Sung, and A. Abraham, “Designing intrusion detection systems: Architectures, challenges and perspectives,” *Studies in Fuzziness and Soft Computing*, vol. **190**, pp. —, 2005

[11] A. Pharate, H. Bhat, V. Shilimkar, and N. A. Mhetre, “Classification of intrusion detection system,” *Int. J. Comput. Appl.*, vol. **118**, no. **23**, pp. **23–26**, 2015

[12] N. Majeed, “A review and classification of intrusion detection system in data engineering,” —, 2021.

[13] R. Wankhede and V. Chole, “Intrusion detection system using classification technique,” *Int. J. Comput. Appl.*, vol. **139**, no. **25**, pp. **25–28**, 2016

[14] S. Niksefat, P. Kaghazgaran, and B. Sadeghiyan, “Privacy issues in intrusion detection systems: A taxonomy, survey and future directions,” *Comput. Sci. Rev.*, vol. **25**, pp. **69–78**, 2017

[15] H. El Zakaria, A. Hafid, and L. Khoukhi, “MiTFed: A privacy-preserving collaborative network attack mitigation framework based on federated learning using SDN and blockchain,” *IEEE Trans. Netw. Sci. Eng.*, vol. **10**, pp. **1985–2001**, 2023, doi: 10.1109/TNSE.2023.3237367.

[16] Q. Lin, R. Ming, K. Zhang, and H. Luo, “Privacy-enhanced intrusion detection and defense for cyber-physical systems: A deep reinforcement learning approach,” *Security Commun. Netw.*, 2022, doi: 10.1155/2022/4996427

[17] S. Chen, Y. Wang, D. Yu, J. Ren, C. Xu, and Y. Zheng, “Privacy-enhanced decentralized federated learning at dynamic edge,” *IEEE Trans. Comput.*, 2023, doi: 10.1109/TC.2023.3239542

[18] “Federated learning with privacy-preserving ensemble attention distillation,” *IEEE Trans. Med. Imaging*, 2023, doi: 10.1109/TMI.2022.3213244

[19] S. R. Spangler, “Privacy-enhancing technologies in federated learning for the Internet of Healthcare Things: A survey,” *Electronics*, 2023, doi: 10.3390/electronics12122703.

[20] A. Elhussein and G. Gursoy, “Privacy-preserving patient clustering for personalized federated learning,” *arXiv preprint arXiv:2307.08847*, 2023.

[21] Y. Wu, C.-F. Chiasserini, F. Malandrino, and M. Levorato, “Enhancing privacy in federated learning via early exit,” in *Proc. ACM*, 2023, doi: 10.1145/3584684.3597274

[22] T. M. Beltrán *et al.*, “Fedstellar: A platform for decentralized federated learning,” *arXiv preprint arXiv:2306.XXXXX*, 2023.

[23] “Federated learning for IoT devices with domain generalization,” *IEEE Internet Things J.*, 2023, doi: 10.1109/JIOT.2023.3234977.

[24] X. Yang, S.-W. Xiang, C. Peng, W. Tan, Z. Li, N. Wu, and Y. Zhou, “Federated learning incentive mechanism design via Shapley value and Pareto optimality,” *Axioms*, vol. **12**, no. **7**, p. **636**, 2023, doi: 10.3390/axioms12070636.

[25] Y. Cui *et al*., “Optimizing training efficiency and cost of hierarchical federated learning in heterogeneous mobile-edge cloud computing,” *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 2022.

[26] J. Zhang, C. Luo, M. Carpenter, and G. Min, “Federated learning for distributed IIoT intrusion detection using transfer approaches,” *IEEE Trans. Ind. Informatics*, 2022

[27] A. Cholakoska, H. Gjoreski, V. Rakovic, D. Denkovski, M. Kalendar, B. Pfitzner, and B. Arnrich, “Federated learning for network intrusion detection in ambient assisted living environments,” *IEEE Internet Comput.*, vol. **27**, pp. **15–22**, 2023, doi: 10.1109/MIC.2023.3264700.

[28] J. Nie, D. Xiao, L. Yang, and W. Wu, “FedCME: Client matching and classifier exchanging to handle data heterogeneity in federated learning,” *arXiv preprint arXiv:2307.08574*, 2023

[29] V. Valadi, X. Qiu, P. Gusmão, N. D. Lane, and M. Alibeigi, “FedVal: Different good or different bad in federated learning,” *arXiv preprint arXiv:2306.04040*, 2023, doi: 10.48550/arXiv.2306.04040.

[30] G. Hu, Y. Teng, N. Wang, and F. R. Yu, “Clustered data sharing for non-IID federated learning over wireless networks,” *arXiv preprint arXiv:2302.10747*, 2023.

[31] J. Li, X. Tong, J. Liu, and L. Cheng, “An efficient federated learning system for network intrusion detection,” *IEEE Syst. J.*, vol. **17**, pp. **2455–2464**, 2023, doi: 10.1109/JSYST.2023.3236995.

[32] M. Nakıp, B. C. Gül, and E. Gelenbe, "Decentralized online federated G-network learning for lightweight intrusion detection," *arXiv preprint arXiv:2306.13029*, 2023.

[33] O. Belarbi, T. Spyridopoulos, E. Anthi, I. Mavromatis, P. Carnelli, and A. Khan, "Federated deep learning for intrusion detection in IoT networks," in *CEUR Workshop Proc.*, vol. **3125**, pp. **85–99**, 2023.

[34] E. M. Campos, P. F. Saura, A. González-Vidal, J. L. Ramos, J. B. Bernabé, G. Baldini, and A. F. Gómez-Skarmeta, "Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges," *arXiv preprint arXiv:2108.00974*, 2021

[35] M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke, and L. Shu, "Federated deep learning for cybersecurity in the Internet of Things: Concepts, applications, and experimental analysis," *IEEE Access*, vol. **9**, pp. —, 2021

[36] M. Alazab, S. P. Rm, M. P., P. K. Maddikunta, T. R. Gadekallu, and V. Q. Pham, "Federated learning for cybersecurity: Concepts, challenges, and future directions," *IEEE Trans. Ind. Informatics*, vol. **18**, no. **5**, pp. **3501–3509**, 2022.

[37] S. Chatterjee and M. K. Hanawal, "Federated learning for intrusion detection in IoT security: A hybrid ensemble approach," *arXiv preprint arXiv:2106.15349*, 2021.

[38] P. Ruzafa-Alcázar, P. Fernández-Saura, E. Mármol-Campos, A. González-Vidal, J. L. Hernández-Ramos, J. Bernal-Bernabe, and A. F. Skarmeta, "Intrusion detection based on privacy-preserving federated learning for the industrial IoT," *IEEE Trans. Ind. Informatics*, vol. **19**, no. **2**, pp. **1145–1154**, 2023.

[39] A. Alazab, A. Khraisat, S. Singh, T. Jan, and M. Alazab, "Enhancing privacy-preserving intrusion detection through federated learning," *Electronics*, 2023

[40] N. A. Al-Marri, B. S. Ciftler, and M. M. Abdallah, "Federated mimic learning for privacy preserving intrusion detection," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, 2020, pp. **1–6**.

[41] W. Yang, B. Liu, C. Lu, and N. Yu, "Privacy preserving on updated parameters in federated learning," in *Proc. ACM Turing Celebration Conf.—China*, 2020.

[42] X. Zhao, L. Wang, L. Wang, and Z. Lu, "A privacy-enhanced federated learning scheme with identity protection," in *Proc. IEEE HPCC/DSS/SmartCity/DependSys*, 2022, pp. **1188–1195**.

[43] A. Elhussein and G. Gursoy, "Privacy-preserving patient clustering for personalized federated learning," *arXiv preprint arXiv:2307.08847*, 2023.

[44] L. Zhang and H. Zhang, "Privacy-preserving federated learning on lattice quantization," *Int. J. Wavelets, Multiresolution Inf. Process.*, 2023, doi: 10.1142/S0219691323500200.

[45] P. Ruzafa-Alcázar *et al*., "Intrusion detection based on privacy-preserving federated learning for the industrial IoT," *IEEE Trans. Ind. Informatics*, vol. **19**, no. **2**, pp. **1145–1154**, 2021.

[46] Y. Liu, G. Wu, W. Zhang, and J. Li, "Federated learning-based intrusion detection on non-IID data," in *Lect. Notes Comput. Sci.*, pp. **313–329**, 2023, doi: 10.1007/978-3-031-22677-9_17.

[47] O. Belarbi *et al*., "Federated deep learning for intrusion detection in IoT networks," *arXiv preprint arXiv:2306.02715*, 2023.

[48] "Federated learning for IoMT applications: A standardization and benchmarking framework of intrusion detection systems," *IEEE J. Biomed. Health Informatics*, 2023, doi: 10.1109/JBHI.2022.3167256

[49] H. Saadat, A. Aboumadi, A. Mohamed, A. Erbad, and M. Guizani, "Hierarchical federated learning for collaborative IDS in IoT applications," in *Proc. MECO*, 2021, pp. **1–6**.

[50] R. Lazzarini, H. Tianfield, and V. Charissis, "Federated learning for IoT intrusion detection," *AI*, vol. **4**, no. **3**, pp. **509–530**, 2023.

[51] Q. Tong, G. Liang, and J. Bi, "Effective federated adaptive gradient methods with non-IID decentralized data," *arXiv preprint arXiv:2009.06557*, 2020.

[52] E. M. Campos *et al*., "Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges," *arXiv preprint arXiv:2108.00974*, 2021.

[53] K. Chen, X. Zhang, X. Zhou, Y. Xiao, and L. Zhou, "Privacy preserving federated learning for full heterogeneity," *ISA Trans.*, 2023, doi: 10.1016/j.isatra.2023.04.020.

[54] A. David, B. Bierbrauer, and N. D. Bastian, "Data-efficient federated learning for raw network traffic detection," *Proc. SPIE*, vol. **12538**, 2023, doi: 10.1117/12.2663092.

[55] "Federated learning for IoMT applications: A standardization and benchmarking framework of intrusion detection systems," *IEEE J. Biomed. Health Informatics*, 2023, doi: 10.1109/JBHI.2022.3167256.

[56] M. M. Rashid *et al*., "A federated learning-based approach for improving intrusion detection in industrial Internet of Things networks," *Network*, 2023, doi: 10.3390/network3010008.

[57] F. Marulli, L. Verde, S. Marrone, R. Barone, and M. S. Biase, "Evaluating efficiency and effectiveness of federated learning approaches in knowledge extraction tasks," in *Proc. IJCNN*, 2021, pp. **1–6**.

[58] V. Valadi *et al*., "FedVal: Different good or different bad in federated learning," *arXiv preprint arXiv:2306.04040*, 2023.

[59] W. Song and T. Yan, "Federated learning framework for blockchain based on second-order precision," in *Proc. IEEE BigComp*, 2023, doi: 10.1109/BigComp57234.2023.00054.

[60] M. Wang *et al*., "TrFedDis: Trusted federated disentangling network for non-IID domain feature," *arXiv preprint arXiv:2301.12798*, 2023.

[61] A. D. Chowdary *et al*., "An ensemble multi-view federated learning intrusion detection for IoT," *IEEE Access*, vol. **9**, pp. **117734–117745**, 2021.

[62] R. Zhao *et al*., "Semi-supervised federated learning based intrusion detection method for Internet of Things," *IEEE Internet Things J.*, 2022.

[63] H. Liang, D. Liu, X. Zeng, and C. Ye, "An intrusion detection method for advanced metering infrastructure system based on federated learning," *J. Mod. Power Syst. Clean Energy*, vol. **11**, no. **3**, pp. **927–937**, 2023.

[64] J. Zhang, C. Luo, M. Carpenter, and G. Min, "Federated learning for distributed IIoT intrusion detection using transfer approaches," *IEEE Trans. Ind. Informatics*, 2022.

[65] J. Zhang *et al*., "Federated learning for distributed IIoT intrusion detection using transfer approaches," *IEEE Trans. Ind. Informatics*, 2023, doi: 10.1109/TII.2022.3216575.

[66] O. Belarbi *et al*., "Federated deep learning for intrusion detection in IoT networks," *arXiv preprint arXiv:2306.02715*, 2023.
[67] P. Li, "FedSD: A new federated learning structure used in non-IID data," in *Proc. IEEE ICASSP*, 2023, doi: 10.1109/ICASSP49357.2023.10095595.
[68] Z. Wang *et al*., "Poisoning-assisted property inference attack against federated learning," *IEEE Trans. Dependable Secure Comput.*, 2023, doi: 10.1109/TDSC.2022.3196646.
[69] G. Yan *et al*., "DeFL: Defending against model poisoning attacks in federated learning via critical learning periods awareness," in *Proc. AAAI Conf. Artif. Intell.*, 2023, doi: 10.1609/aaai.v37i9.26271.
[70] P. R. Ovi *et al*., "Confident federated learning to tackle label-flipped data poisoning attacks," *Proc. SPIE*, 2023, doi: 10.1117/12.2663911.
[71] L. Lavaur *et al*., "The evolution of federated learning-based intrusion detection and mitigation: A survey," *IEEE Trans. Netw. Serv. Manag.*, 2022, doi: 10.1109/TNSM.2022.3177512.
[72] X. Wu *et al*., "Faster adaptive federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. **37**, no. **9**, pp. **10379–10387**, 2023.
[73] J. Mills *et al*., "Accelerating federated learning with a global biased optimiser," *IEEE Trans. Comput.*, 2022.
[74] Y. Rahulamathavan *et al*., "FheFL: Fully homomorphic encryption friendly privacy-preserving federated learning with Byzantine users," *arXiv preprint arXiv:2306.05112*, 2023.
[75] W. Mou *et al*., "A verifiable federated learning scheme based on secure multi-party computation," in *Lect. Notes Comput. Sci.*, 2021, doi: 10.1007/978-3-030-86130-8_16.
[76] R. Subedi *et al*., "A client-server deep federated learning for cross-domain surgical image segmentation," *arXiv preprint arXiv:2306.08720*, 2023.
[77] W. Huang *et al*., "FedCKE: Cross-domain knowledge graph embedding in federated learning," *IEEE Trans. Big Data*, 2022.
[78] S. Liu and F. Xu, "Adaptive federated learning aggregation strategies based on mobile edge computing," in *Proc. ICMLCA*, vol. **12636**, pp. **65–73**, SPIE, 2023.
[79] A. Selamnia *et al*., "Edge computing-enabled intrusion detection for C-V2X networks using federated learning," in *Proc. IEEE GLOBECOM*, 2022, pp. **2080–2085**.
[80] R. Yu and P. Li, "Toward resource-efficient federated learning in mobile edge computing," IEEE Netw., vol. 35, pp. 148–155, 2021.