ISSN 2521-6376

Journal of Baku Engineering University

MECHANICAL AND INDUSTRIAL ENGINEERING

Volume 9
Number 2
2025

Journal is published twice a year
Number-1. June, Number-2. December

An International Journal

http://journal.beu.edu.az

Contact address

Journal of Baku Engineering University
AZ0102, Khirdalan city, Hasan Aliyev str. 120, Absheron, Baku, Azerbaijan

Tel: *00 994 12 - 349 99 66/78* ***Fax:*** *00 994 12 349-99-90/91*

e-mail: *journal@beu.edu.az*
web: *http://journal.beu.edu.az*
facebook: *Journal Of Baku Engineering University*

ISSN 2521-6376

ISSN 2521-6376

BMU
BAKI MÜHƏNDİSLİK UNİVERSİTETİ

Journal of Baku Engineering University

MECHANICAL AND INDUSTRIAL ENGINEERING

Baku - AZERBAIJAN

Journal of Baku Engineering University

MECHANICAL AND INDUSTRIAL ENGINEERING

2025. Volume 9, Number 2

CONTENTS

2025. Volume 9, Number 2 Pages 75-82

UDC:624.042.7
DOI: https://doi.org/10.30546/09085.2025.02.308

EXPERIMENTAL STUDY OF THE TENSION-DEFORMATION STATE OF THE TIGHTENING COMPONENT

Osman MIRZAYEV[1*], Huseyn MIRZAYEV[2], Vurgun GAHRAMANOV[1]

[1]*Baku Engineering university, Khirdalan, Azerbaydjan*

[2]*Karabakh university, Khankendi, Azerbaydjan*

ARTICLE INFO

Article history
Received:2025-07-04
Received in revised form:2025-07-17
Accepted:2025-10-04
Available online

Keywords:
Oil field equipment;
gas lift valves;
deformation;
compression;
tightening.

JEL Classification:
TA350, TA405, QA808, TJ840

ABSTRACT

Gas lift equipment has a special place in oil extraction. This method has an important place among mechanized methods. This method is used to increase the longevity of the wells, and at the same time, when the productivity of the wells decreases, to increase the pressure at the bottom of the well by supplying gas to the space behind the well, to raise oil to the surface.

Gas lift valves are one of the main components of the equipment complex used in wells using the gas lift method. Depending on the characteristics of the wells, it is required to meet complex requirements from the gas lift valves, among which the provision of tightening is one of the important problems. In order to ensure effective tightening in gas lift valves, it is of great importance to choose the construction, dimensions of tightening elements, as well as the criteria of tightening correctly.

From this point of view, the solution of scientific-practical problems related to the design and calculation methodology of tighteners that can create reliable tightening in gas lift valves can be considered as one of the actual problems of production and exploitation of modern oil field equipment.

1. Introduction

The study of the tensıon-deformation state is one of the important problems in providing reliable hermeticity at high pressures in the gas lift equipment's tightening component. Theoretical studies lay the foundation for solving of this problem by accepting certain approximations, but in order to prove the theoretical results, there is a need to accept the results of experimental research.

A complicated tensıon-deformation state is created in the tightening component between the tightening elements, as well as in their touch with the contact surface (body).

At this stage of the research works, the tensıon-deformation state was studied by an experimental method. For this purpose, the following have been determined - variable pressure ΔP acting on the head surface of the tightener;

- axial force distributed on the tightening contact surface – Qt,

- the total tighting force applied to the tightening component.

The following approximations were accepted when solving the problem:

- the height of the tightened cylinder is the same;

- local radial and axial loads are equally distributed on the tightening surface. In this case, the axial load is replaced by its equivalent static tangential force.

*Corresponding author.

E-mail addresses: omirzayev@beu.edu.az (Mirzayev Osman Hasan).

2. PROBLEM STATEMENT

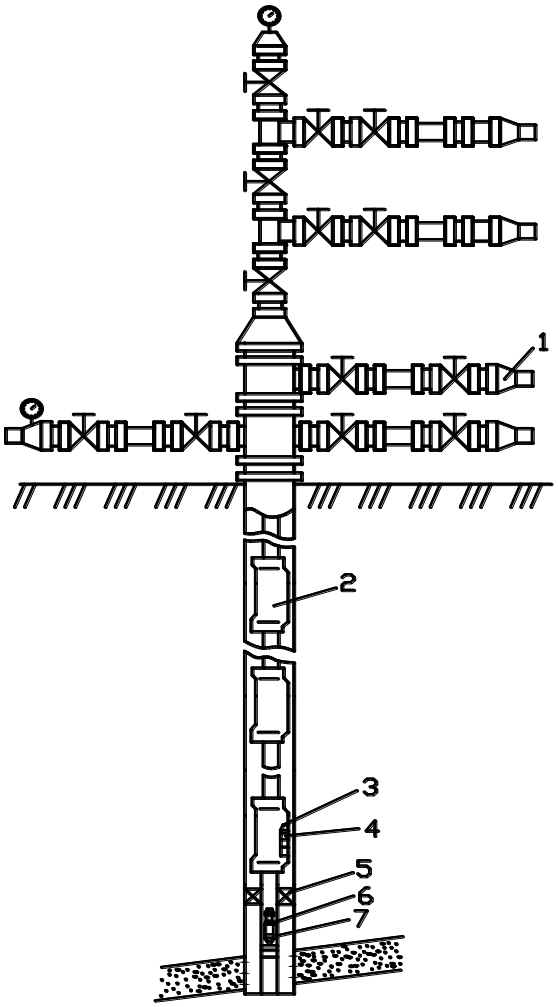


Fig. 1 1.1. General view of the gas lift well:

1- fountain armature, 2- well chamber, 3- punched fastener, 4 - gas lift valve, 5- packer, 6- intake valve, 7- nipple

Experiments were carried out on gas lift valves with both existing and offered tightening components. Constructions of tightening components are shown in Fig. 2 a and b, respectively.

The tightening component of the offered gas lift valve differs from the existing structure in that, in order to reduce relaxation tensions in the rubber cuffs, lead rings are placed in the sockets of the tightening cuffs [3].

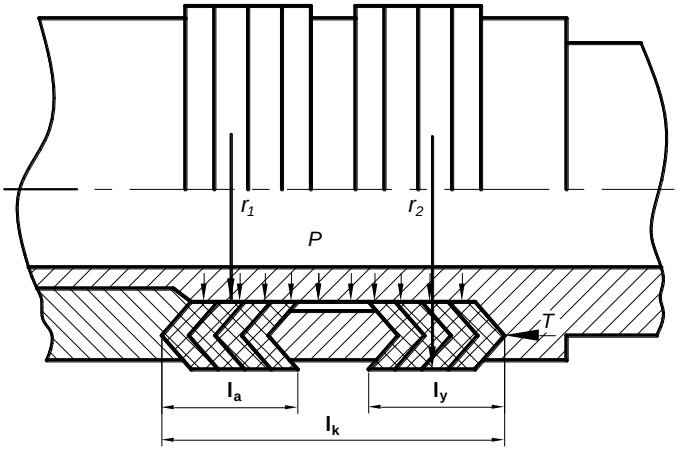


Fig. 2 a. An existing tightening component

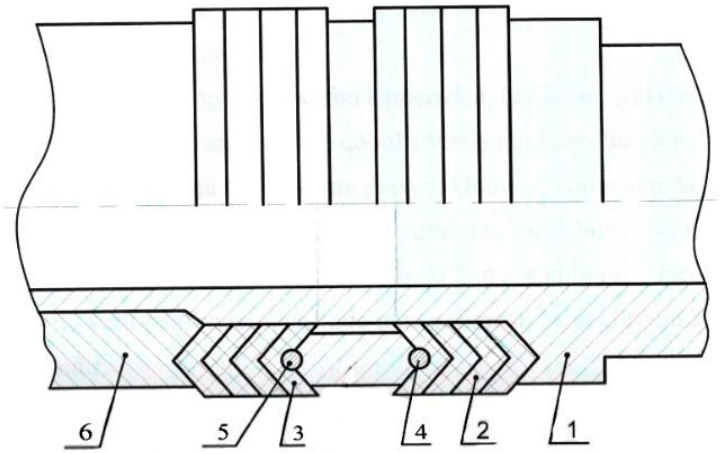


Fig. 2 b. A tightening component of the new construction

1 – a cover, 2 and 3 – rubber cuffs, 4 and 5 – lead rings, 6 – a silphon body

It is known that the tightening component of gas lift valves consists of two sets of oppositely directed tightening elements with a conical profile. For this reason, when the tighteners are placed in the chamber (cylinder) of the gas lift valve, loading of tightening has a great role.

Table 1 shows the location state and the number of the tighteners.

Table 1. The location state of the tighteners while tightening

No. of experiments	The location state of the tighteners involved in tightening	The number of the tighteners involved in the contact
1	The I set of the tightening elements meets the contact surface	3
2	The middle surface of the tightening elements is in contact	2
3	The I set and the II set of the tighteners are in contact	6

The radial deformation diagram of the displacement is given in figure 5.5. It is seen from the figure that the dependence z/r_2 of the displacement in the tighteners on dimensionless relative axial coordinate varies with the exponential law. From this one can conclude that the radial displacement is not equally distributed on the axial height in the set of tighteners.

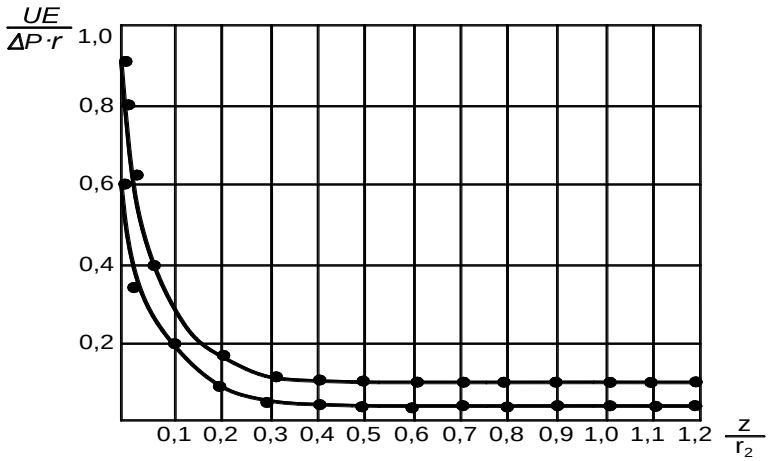


Fig. 3. The radial deformation of the displacement in the tightener.

The change diagram of pressure-dependent tensions in the tighteners is given in figure 4 and the change diagram of the force tangent to the contact line in them is given in figure 5.

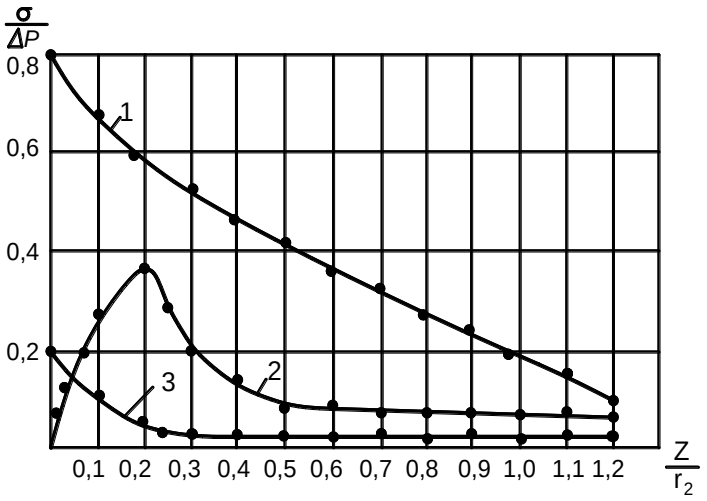


Fig. 4. The change diagram of pressure-dependent tensions in the tighteners with D=32 mm:

1-σ_z (r=r1); 2-σ_z (r=r1); 3-σ_z (r=r2)

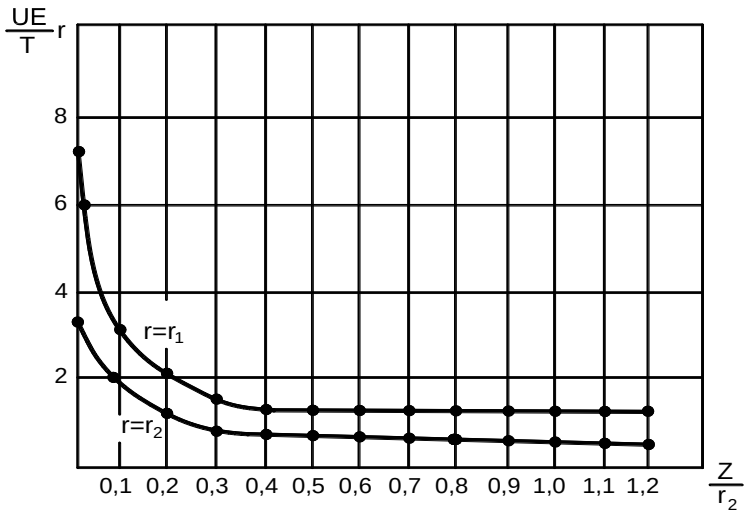


Fig. 5. The change diagram of the force tangent to the tightening contact surface

The dependence of the axial tension on dimensionless relative axial coordinates is shown in figure 5.8.

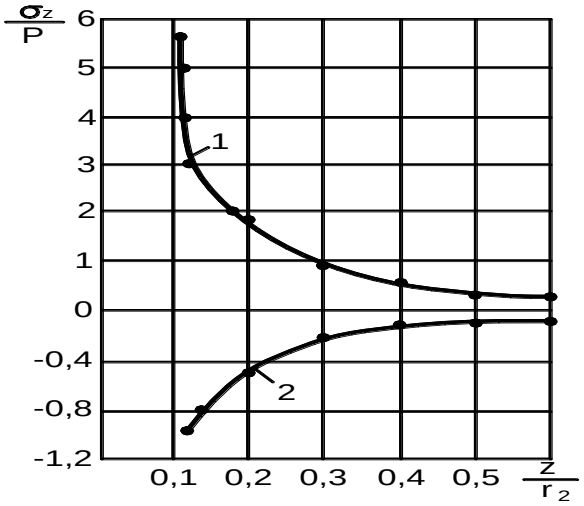


Fig. 6. The dependence of the axial tension on axial coordinate:

ly – the upper (I) set, la – the lower (II) set of tighteners.

$$1-\frac{l_y}{r_2};\ 2-\frac{l_a}{r_2}$$

3. PROBLEM SOLUTION

The II stage of the experiments is devoted to the study of the tension-deformation state created in the tighteners with a new shape due to tighting force and pressure.

In the tests, breaking the hermeticity by providing the stepwise increase of the tension on the contact surfaces and the increase of the pressure between 10-50 MPa, as well as self-tightening were performed. According to each step of tension or pressure, the tension and deformations on the inner and outer contact surfaces of the tightening element were measured.

The results of the experiments for the newly offered tighteners (with outer diameters ϕ29, ϕ32, ϕ40) are given in figure 8-14.

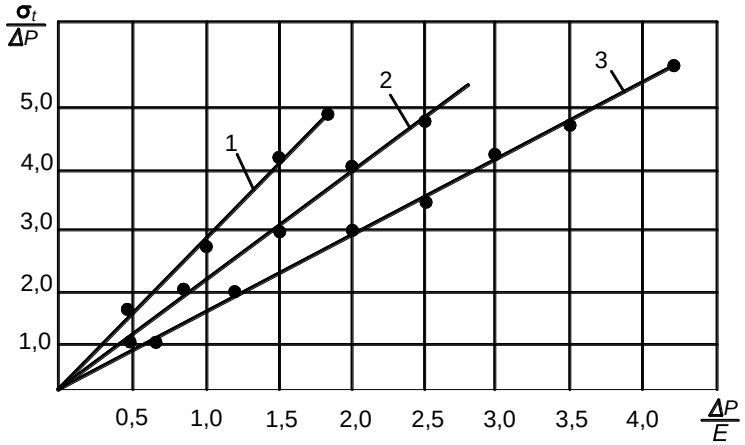


Fig. 7. The change diagram of the axial tension in the new-shaped tighteners:

1-D=29 mm; 2-D=32 mm; 3-D=40 mm

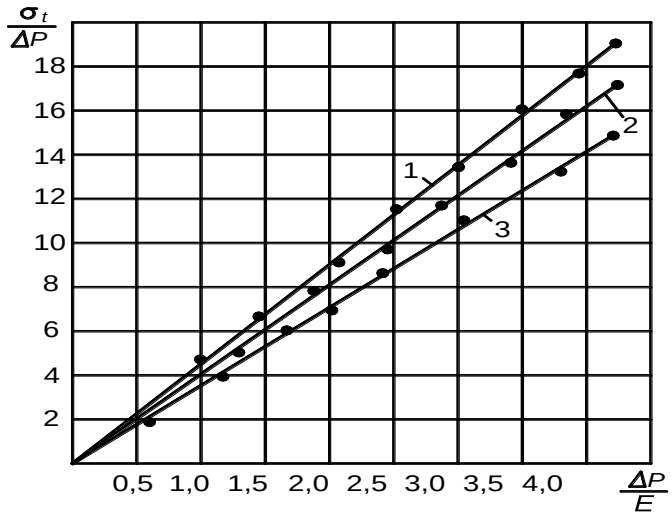


Fig. 8. The change diagram of the tangential tension in the new-shaped tighteners:

1-D=29 mm; 2-D=32 mm; 3-D=40 mm

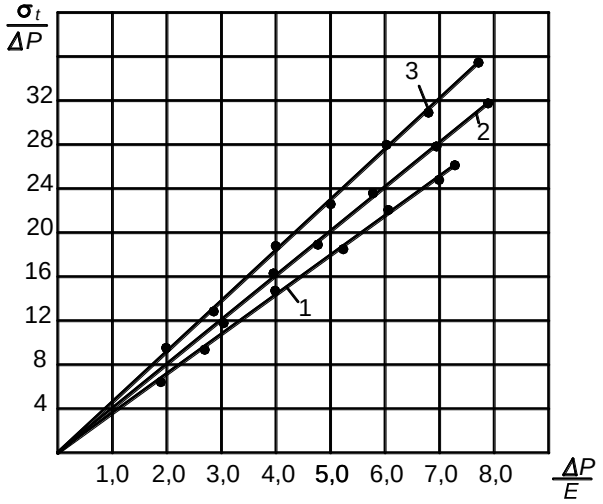


Fig. 9. The change diagram of the tangential tension in the new-shaped tighteners:

1-D=45 mm; 2-D=48 mm; 3-D=50 mm

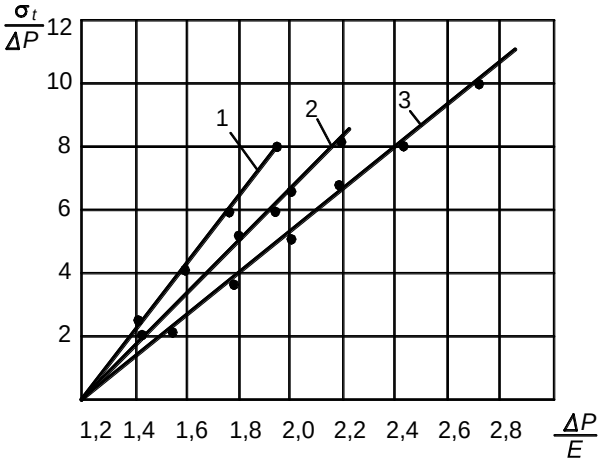


Fig. 10. The change diagram of the axial tension in the new-shaped tighteners:

1-D=45 mm; 2-D=48 mm; 3-D=50 mm;

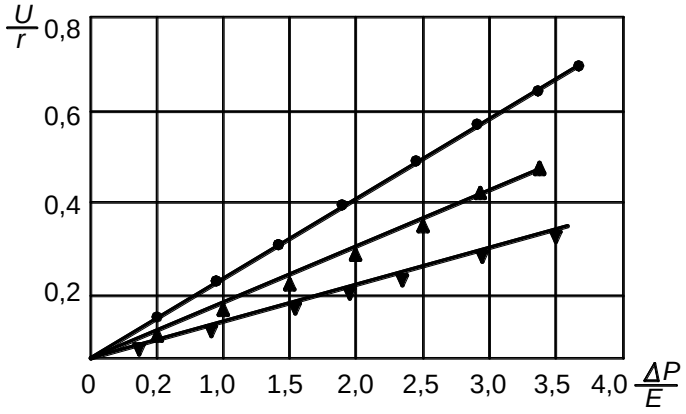


Fig. 11. The dependence of the radial displacement on the radial pressure in the new-shaped tighteners:

▼ -D= 45mm; ▲-D= 48mm; ●-D=50mm.

In existing tighteners (for loading states I and II in table 5.1), the radial displacement is obtained at a value that does not satisfy the tightening limit –U, and such an effect is typical for both tension and pressure application cases.

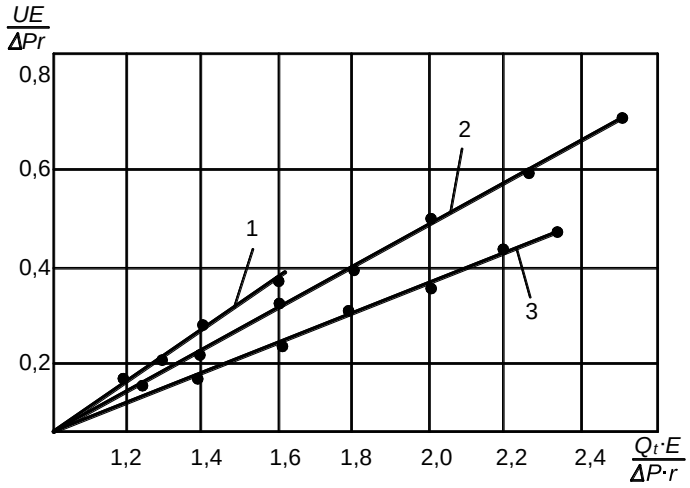


Fig. 12. The dependence of the radial tension on the total tighting force in existing tighteners:

1-D=29 mm; 2-D=32 mm; 3-D=40 mm

Characters of dependences of axial and tangential tensions on local loads Q_t for different loading regimes in the tightener with the size ϕ29 according to table 1 are given in figures 13 and 12.

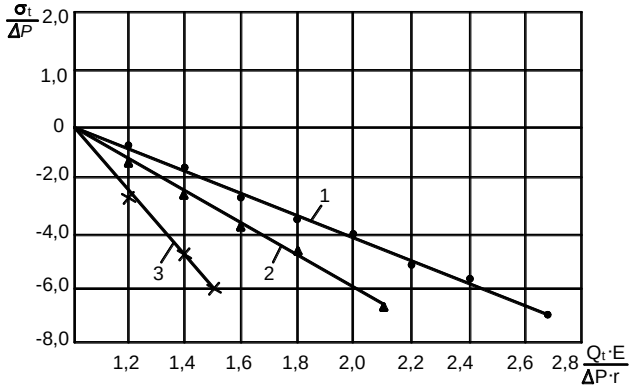


Fig. 13. The dependence of the axial tension on the tighting force: 1-✕; 2-▲; 3-●;

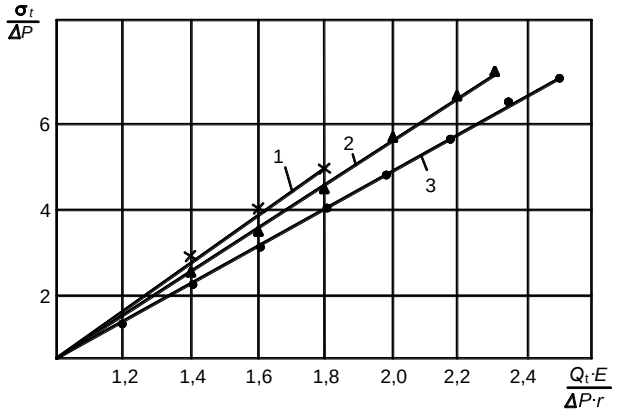


Fig. 14. The dependence of the tangential tension on the tighting force:

1-✕; 2-▲; 3-●;

It is seen from the results of the experiment that the values of the radial displacement and the tangential tension depend more on the radial load, and the axial force reduces their value.

It has been determined by the researches that depending on the loading mode of the tightening elements on the contact surface, the tensions increase with the increase of pressure, but their decrease is observed because of the influence of local loads. Also, since in the new-shaped tightening elements, tightening is obtained by a total contact, the value of the tensions is sufficient so that tightening occurs with equal contact surfaces (with tightening elements both the I set and the II set). In this case, distributions of tensions and the radial displacement satisfy Lame's law in the new-shaped tighteners.

4. CONCLUSIONS

So, an additional elastic element placed in a special slot in the tightening component reduces the effect of local loads, in this case, an equal distribution of the radial tension is observed. At the same time, the values of the radial tension are greater than the existing tightening component, and self-tightening is ensured under the direct and inverse influence of the pressure.

REFERENCES

1. C.C.Azar, Q.Robello Samuel. Drilling engineering. 2016. 520 pg.
2. Mammadov V.T., Mirzayev O.H. Repair and restoration of oil field equipment. Baku, 2015, 385 pp.
3. Vahidov M.A., Karimov O.M. Oil and gas production equipment. Baku, 2009, 440 pp.
4. Baghirova G.S. Determining the temperature endurance time of the tightening component in gas lift valves // Journal of Scientific Papers of ATU, No. 2, 2008, pp. 68-74.
5. Mammadov V.T. The method of choosing the optimal shape and sizes of the tightening components of oil field equipment. Guidance Document RS. 577 3272-027-2001; Baku: 2001.-41pp.

2025. Volume 9, Number 2 Pages 83-89

UDC: 539
DOI: https://doi.org/10.30546/09085.2025.02.310

THE STATIC ANALYSIS OF CYLINDRICAL SHELLS MADE OF COMPOSITE MATERIAL

Natig RZAYEV[1*], Mahir BASHIROV[1], Nofal NABIYEV[1]

*nrzayev@beu.edu.az**

[1]Baku Engineering University, Khirdalan, Azerbaijan

ARTICLE INFO

Article history
Received:2025-09-30
Received in revised form:2025-10-05
Accepted:2025-10-15
Available online

Keywords:
Composite material;
Von Mises stress;
cylindrical shell;
layer;
deformation.
JEL Classification: L69,O31,O33, L23

ABSTRACT

In this work, a static analysis of a cylindrical shell made of glass fiber reinforced composite material was performed. The static analysis was performed using the simulation application in the Solidworks program. For this purpose, a cylindrical cover with a diameter of 50mm and a length of 100mm was designed. The surface command was used in the Solidworks program to design the part. A composite material consisting of three layers reinforced with glass fiber was selected as the material. The boundary conditions and loading cases of the cylindrical shell were applied in accordance with real operating conditions. The distribution of Von Mises stress, normal stresses along the X and Y axes and deformations were analyzed across the layers. As a result of the analysis, both intra-layer and inter-layer stress and deformation distributions were determined, and areas with particularly high stress concentrations were identified.

1. Introduction

In modern engineering, the demand for lightweight, high-strength and corrosion-resistant structures is increasing. Among them, cylindrical shells made of composite materials dominate. This type of cylindrical shells is widely used in aerospace, shipbuilding, automotive industry, energy and pipeline infrastructure. Their mechanical behavior depends on the homogeneity and heterogeneity of the material, the asymmetry of its structure and the design of the cylindrical shell, the applied loads and the placement conditions. Numerical simulation plays an important role for this. Here it is possible to evaluate the load-bearing capacity of cylindrical shells, optimize the design and identify their weak points.

The mechanical behavior of composite cylindrical shells has been studied in many scientific works. In [1], [2], the methods for calculating the elasticity constants of parts made of composite materials were given and the stability theories of cylindrical shells were improved. The effect of the directions of lifts and the sequence of layers in composite materials on the mechanical behavior was studied in [3]. In [5], stress analyses were performed on structural parts made of composite materials using the finite element method. The results of this analysis were shown to be consistent with both analytical and experimental results. The problem of crack formation in parts made of composite materials was studied in [6].

*Corresponding author.

E-mail addresses: nzayev@beu.edu.az (Rzayev Natig Samandar).

In general, the literature review shows that for an accurate assessment of the behavior of composite cylindrical shells, it is necessary to take into account the combined effects of the properties of the layers, fiber orientations, boundary conditions and loading type. The application of modern programs such as SolidWorks Simulation allows for a comprehensive analysis of these parameters and allows the results to be correlated with real engineering applications.

In this article, the static behavior of a three-layer glass fiber composite cylindrical shells is investigated using SolidWorks Simulation. The main focus is on the analysis of stress and strain distribution, differences in loading across layers and identification of potential weak areas. The aim of the article is to evaluate the behavior of composite shells from both theoretical and applied perspectives and to provide a basis for future optimization work.

2. Materials and Methods

2.1. Cylindrical shells design and material selection

The mechanical properties of the material are as follows:

E-fiber glass:

Density: 2,58g cm^{-3};

Tensile strength: 3,445Gpa;

Youngs modulus: 72,3Gpa;

Elongation: 4,8%

Poissons ratio: 0.2.

Test specimen dimensions: Outer diameter 25mm; length 100mm;

Number of layers -3. All layers are assumed to be of the same material. The angle for the first layer is 150, for the second layer 300 and for the third layer 450.

Applied force: 1000N. The sample is fixed on one side. A force of 1000N is applied across its surface.

Mesh sizes: 2.5mm and 0.125mm.

Proqram: Solidworks (simulation) .

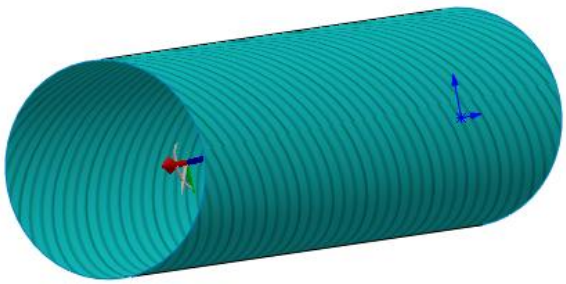


Fig. 1. *3D model of a cylindrical shell made of composite material.*

A cylindrical part made of a composite material consisting of three layers was designed. The thickness of each layer was chosen to be 0.5mm. All layers were assumed to be made of the same

material. The angle for the first layer was assumed to be 150, for the second layer 300, and for the third layer 450. In total, the angles for the layers changed in 150-degree intervals (Figure 1).

2.2. Static analysis of a cylindrical shell

After applying the material properties to the 3D model in layers, boundary conditions must be accepted. For this purpose, a cylindrical part made of composite material is fixed along its two or three surfaces along the circumference (fixed geometry fastening type is selected). A force of 1000N is applied along the surface of the sample. The force is distributed at equal values along the surface of the cylindrical part (Figure 2).

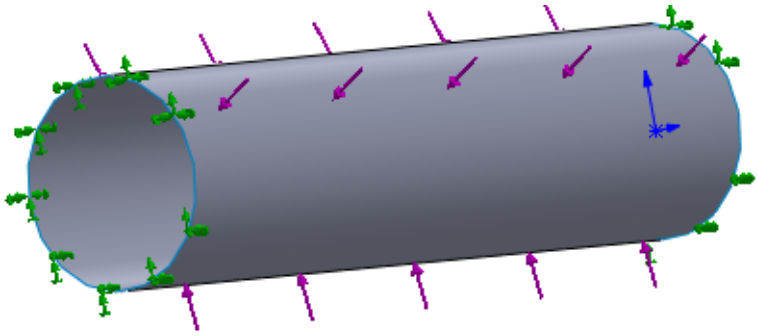


Fig. 2. *Application of boundary conditions.*

To determine the distribution of stress and strain along the layers of the part, a mesh property must be applied. The size of the mesh parameters affects the accuracy of the results. For this purpose, mesh parameters with sizes of 2.5mm and 0.125mm along the layers of the composite part were adopted (Figure 3).

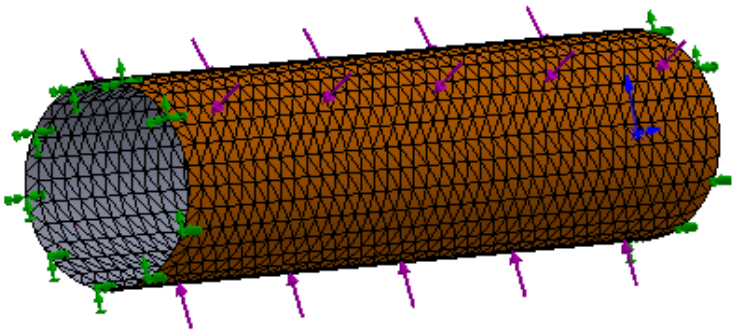


Fig. 3. *Applying mesh parameters to a composite part by layers.*

Analysis of stresses in the first layer of the composite part. First, the analysis of the Von Mises stress was carried out. The minimum value of this stress was observed at points close to the hardening zone of the sample, and the maximum value was observed in the linear direction in the middle zones. The maximum and minimum values of the stress were obtained as 1.012e+00 Mpa and 3.278e-01MPa, respectively. The results of the linear simulation analysis of the Von Mises stress show that the maximum values of the stress were distributed in the linear direction outside the hardening zone (Fig. 4 a). The maximum value of the stress along the X axis was 9.016e-03Mpa, and the minimum value was -1.116e+00 MPa. The maximum value of the stress along the Y axis was 8.581e-02Mpa, and the minimum value was -1.099e+00. The distributed values of the stresses along the X and Y axes are given in Fig. 4, b, c.

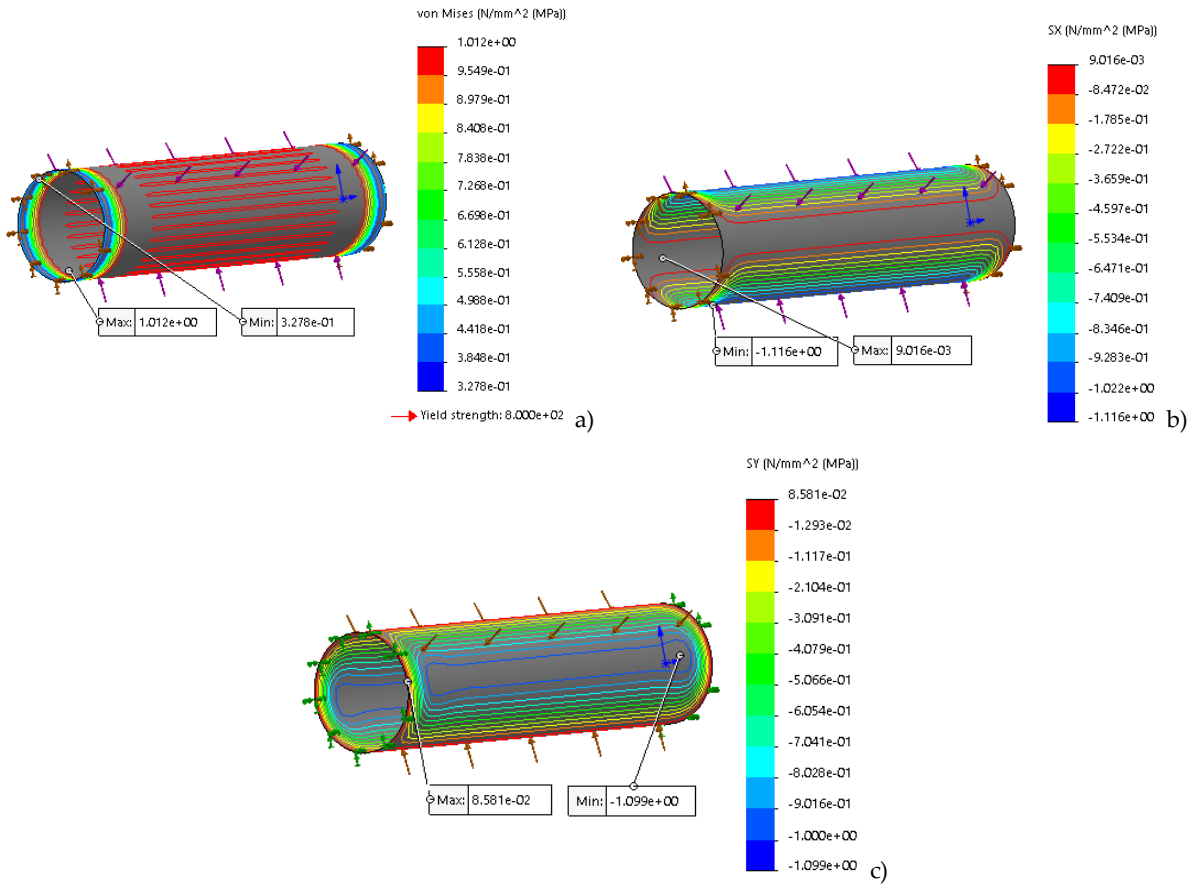
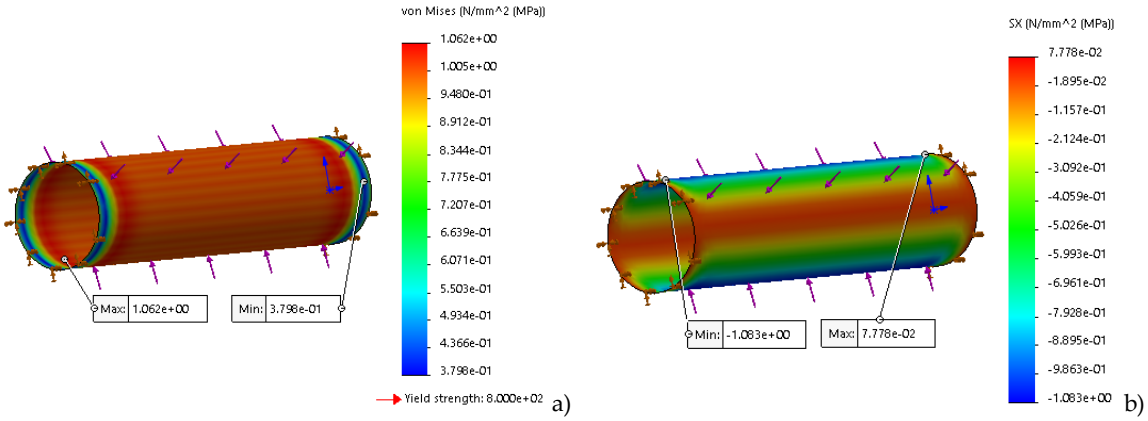


Figure 4. *Stress distribution across the first layer of the composite part.*

Stress distribution in the second layer of the composite part: The maximum value of the Von Mises stress in the second layer was 1.062e+00Mpa and the minimum value was 3.798e-01MPa. The maximum value of the stress was distributed linearly along the two middle surfaces where the layer was located, in the direction of the fibers. The minimum value of the stress was observed in the reinforcement zones. From this, a neutral zone was formed along the circumference between the stress zones (Figure 5 a).

The maximum values of the stresses along the X and Y axes in the second layer were 7.778e-02MPa and -7.043e-04MPa, respectively; the minimum values were 1.083+00 MPa and -1.132e+00 MPa. The maximum and minimum values of the stress along the X axis were formed in symmetrical sections along the length and were observed near the reinforcement zones (Figure 5 b, c).



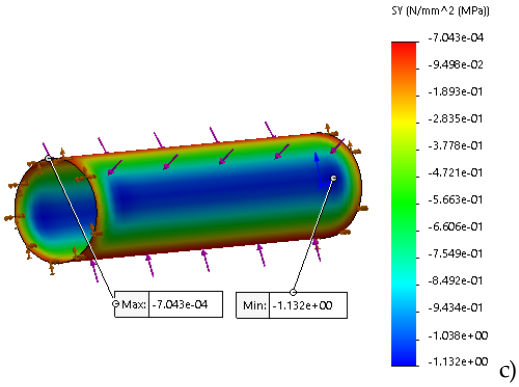


Figure 5. *Stress distribution in the second layer of the composite part.*

Stress distribution in the third layer of the composite part: The maximum value of the Von Mises stress in the third layer was observed along the circumference in the reinforcement zone. This value is 1.668e+00 MPa, respectively. The minimum value of this stress was 3.931e-01, and it was distributed in a certain width band along the circumference, partially outside the reinforcement zone (Fig. 6 a). The maximum and minimum values of the stresses along the X axis are observed to be distributed in reciprocal segments along the longitudinal section. The values of these stresses were obtained as 9.506e-03MPa and -1.153e+00MPa, respectively (Fig. 6 b).

The distribution zone of the maximum value of the stress along the Y axis is relatively small. This stress is observed in small segments along the axis of the cylinder. The distribution zone of the minimum values is relatively large. The minimum stress is observed along the axis of the cylinder. The minimum and maximum values of the stresses are -1.169e+00 MPa and -1.218e-03 MPa, respectively (Figure 6 c).

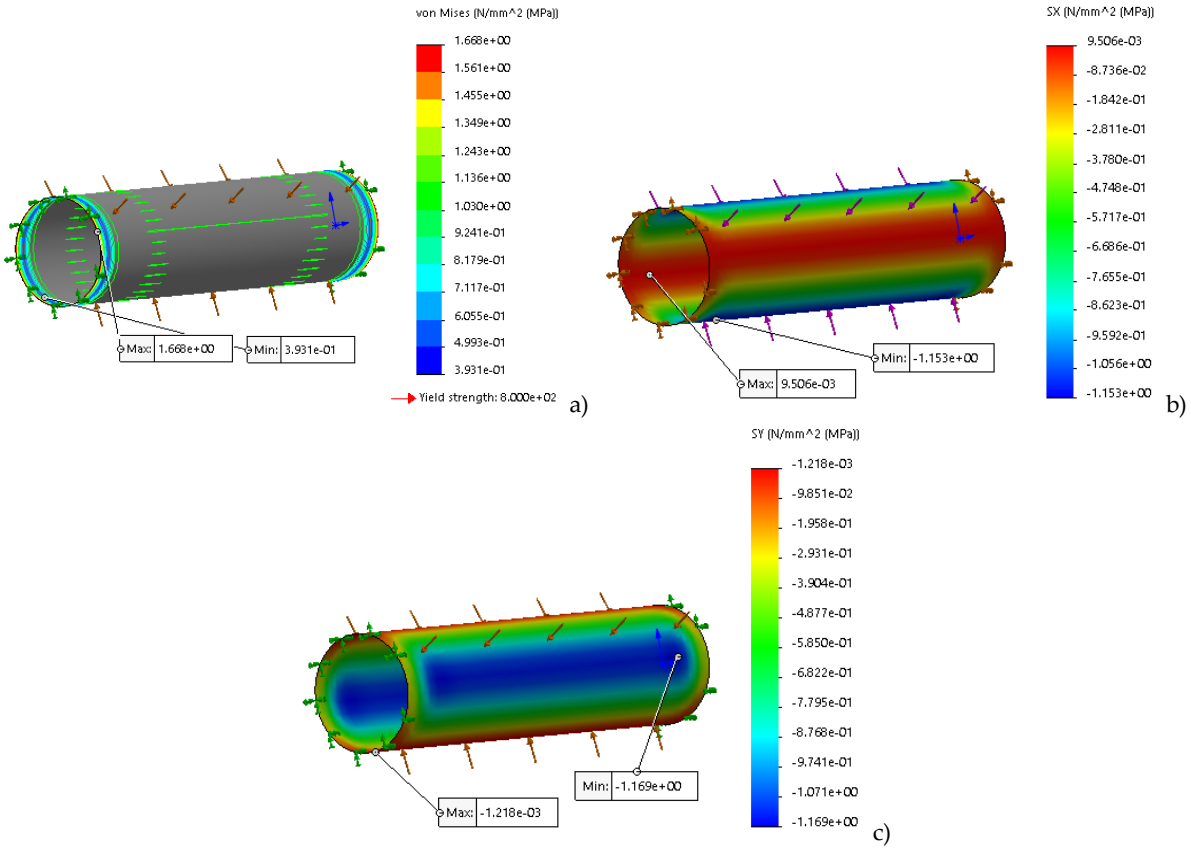


Figure 6. Stress distribution across the third layer of the composite part.

Distribution of deformation along the layers of the composite part: The minimum value of deformation for all three layers is distributed in the zones close to the fastening surface, while the maximum values are distributed along the circumference in the fastening zones, partially on the outer surfaces. The numerical value of deformation has the smallest value in the fastening zones.

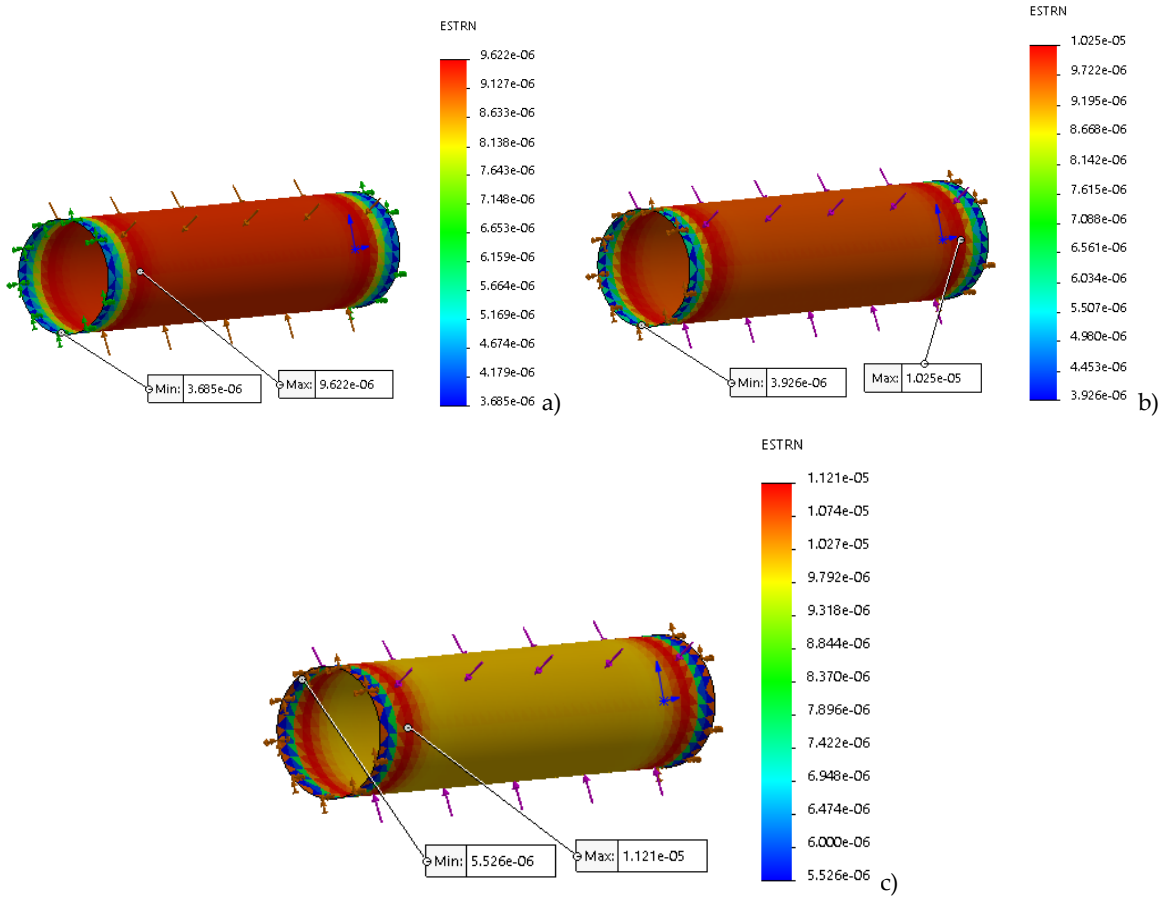


Figure 7. *Distribution of deformation in the layers of a composite part.*

As it moves away from the middle zone, the deformation increases, reaches a maximum value, and then decreases sharply. The maximum value of deformation for the first layer is 9.622e-06, the minimum value is 3.68e-06; in the second layer, 1.025e-05 and 3.926-06, respectively, and in the third layer, 1.121e-05 and 5.526-06 were obtained (Fig. 8).

3. Conclusion and discussions

In the research work conducted, the static behavior of a cylindrical shell reinforced with three layers of glass fibers was analyzed using the finite element method in the SolidWorks Simulation environment, and a number of important scientific and practical results were obtained.

1. The analyses showed that the stress distribution is not homogeneous throughout the cylindrical shell, but is concentrated more in the boundary areas and interlayer zones. This is especially noticeable when the elastic moduli of the layers and fiber directions are different. Local stress concentration was assessed as the main factor increasing the probability of damage to the cylindrical shell.
2. The ply arrangement and fiber orientation in a composite cylindrical shell have a significant impact on the overall strength and deformation behavior of the structure. While

circumferential fiber arrangement increases the resistance to internal pressure, axial fiber arrangement increases the deformation stiffness. Thus, the optimal ply selection should be determined according to the functional requirements of the structure.

3. Analysis of strain distributions showed that the three-layer composite structures exhibited a nonlinear mechanical response compared to simple isotropic cylinders. The mechanical properties and orientation angles of each layer changed the overall elastic properties, leading to the formation of both maximum displacements and local deformations.
4. Modeling based on SolidWorks Simulation software allowed to predict the static behavior of the composite cylindrical coatings with high accuracy.

The results of the study show that not only the choice of material, but also the sequence and direction of the layers should be taken into account when designing composite cylindrical shells. This approach allows to increase the reliability of the structure during operation, prevent excessive stresses and ensure long-term durability. Simulation-based approaches in this direction can contribute to achieving more optimal, reliable and economical design solutions in both production and scientific research processes in the future.

REFERENCES

[1]. Jones, R. M. Mechanics of Composite Materials. 2nd Edition, p538, 1999.

[2]. Reddy, J. N. Mechanics of Laminated Composite Plates and Shells: Theory and Analysis. CRC Pres, p838, 2003.

[3]. Daniel, I. M., & Ishai, O. Engineering Mechanics of Composite Materials. Oxford University Press, p403, 2006.

[4]. Hyer, R. E. Stress Analysis of Fiber-Reinforced Composite Materials. McGraw-Hill, p687, 1998.

[5]. Zhang, Y., Li, X., Wang, J. "Finite element analysis of laminated composite cylindrical shells under internal pressure." Composite Structures, Vol. 140, pp. 67–75, 2016.

[6]. Hashin, Z. "Failure criteria for unidirectional fiber composites." Journal of Applied Mechanics, Vol 47, no 2, pp.329–334, 1980.

UDC: 004
DOI: https://doi.org/10.30546/09085.2025.02.313

ANALYSIS OF MIGRAINE AURA BASED ON EEG TESTS USING ARTIFICIAL INTELLIGENCE METHODS

Javid ABBASLI[1*]

[1]*Azerbaijan Technical University, Baku, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-01
Received in revised form:2025-10-05
Accepted:2025-10-15
Available online

Keywords:
EEG Tests;
fully connected neural networks; artificial intelligence;
FitzHugh-Nagumo model;
Milstein method.

JEL classification: I10,I12,C45,C55

ABSTRACT

The dynamics of the membrane potential and recovery mechanism of biological neurons during migraine attacks with aura were mathematically modeled using the FitzHugh-Nagumo nonlinear stochastic differential equation system. The degree of influence of the membrane potential on the recovery mechanism, the degree of self-regulation of the recovery mechanism, and the stochastic resonance intensity coefficients affecting both components were determined using a fully connected neural network. This study addresses a significant gap in computational neuroscience by integrating stochastic differential equations with machine learning to characterize neuronal behavior during pathological conditions. Traditional electroencephalography (EEG) analysis methods primarily rely on time-frequency decomposition and statistical techniques, identifying fundamental signal characteristics but remaining disconnected from mechanistic stochastic neuronal models. Our approach combines the biophysically realistic FitzHugh-Nagumo framework with neural network-based parameter estimation using EEG recordings, enabling precise quantification of key biophysical parameters governing neuronal excitability and recovery processes. The identified parameters provide quantitative measures of membrane dynamics and stochastic fluctuations characterizing migraine pathophysiology, offering potential biomarkers for clinical diagnosis and personalized treatment strategies.

1. Introduction

Electroencephalography (EEG) tests are essential neurological diagnostic methods used to record and analyze brain wave activity (Niedermeyer & da Silva, 2004). EEG plays an important role in diagnosing various neurological conditions by recording the synchronized activity of cortical neurons, particularly valuable for studying rapid neurophysiological changes in disorders such as migraine with aura (Wolpaw et al., 2002).

EEG signals consist of waves with varying amplitudes and frequencies, divided into five types: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (>30 Hz) (Steriade et al., 1993). Each wave type is associated with specific cognitive and physiological states and reflects distinct patterns of neuronal synchronization across cortical networks (Buzsáki & Draguhn, 2004). Delta waves occur during deep sleep and are characterized by high amplitude and low frequency, representing widespread synchronization of cortical neurons.

*Corresponding author.

E-mail addresses: cavid.abbasli@aztu.edu.az (Abbasli Javid Yashar).

Theta waves relate to emotional states (Deco & Jirsa, 2012), memory encoding, and creative processes, particularly prominent in the hippocampus and temporal regions.Alpha waves predominate in relaxed but awake states (Klimesch, 1999), typically observed when individuals are at rest with eyes closed, and are most prominent in posterior brain regions. Beta waves are associated with active thinking, focus, and motor activity, reflecting desynchronized cortical states during cognitive engagement. Gamma waves represent the highest frequency oscillations and are linked to sensory processing, attention, and consciousness, often considered markers of neural integration across distributed brain networks (Buzsáki, 2006). Understanding these patterns is crucial for identifying abnormal neuronal activities during migraine aura episodes, as pathological states often manifest as alterations in the power, coherence, and spatial distribution of these frequency bands. Migraine with aura involves cortical spreading depression (CSD), characterized by neuronal depolarization waves followed by activity suppression propagating across the cortex at approximately 2-5 mm/min, which produces characteristic changes in EEG patterns during aura phases.The FitzHugh-Nagumo model (FitzHugh, 1961; Nagumo et al., 1962), a simplified version of the Hodgkin-Huxley model, effectively describes neuronal dynamics through differential equations. Its stochastic extension captures inherent biological randomness, making it well-suited for modeling pathological conditions like migraine aura (Gardiner, 2004; Wiesenfeld & Moss, 1995). Despite advances in mathematical neuroscience, existing approaches have limitations. Izhikevich (2007) contributed significantly to modeling neuronal dynamics but did not deeply address stochastic effects and parametric analysis of EEG data. Traditional EEG analysis methods using time-frequency techniques (Delorme & Makeig, 2004; Lotte et al., 2007) revealed signal characteristics but lacked integration with stochastic neuronal models. Deco and colleagues (Deco & Jirsa, 2012; Deco et al., 2008) studied stochastic processes in neural models but did not apply artificial neural networks for systematic parameter estimation from clinical data.

This work addresses these gaps by integrating machine learning with stochastic differential equations. We employ a multi-layer perceptron (MLP) neural network with ADAM optimization (Kingma & Ba, 2015) to estimate five key biophysical parameters from EEG data: internal self-regulation (a), membrane potential influence (b), neighbor influence (c), and stochastic resonance intensities (σ_R, σ_P). Using 1230 EEG recordings from migraine patients, we validate our approach and solve the stochastic system using the Milstein numerical method.

This study contributes to clinical neuroscience by providing quantitative biomarkers for migraine aura, enabling early detection and personalized treatment strategies. The paper is organized as follows: Section 2 reviews related work, Section 3 presents the stochastic FitzHugh-Nagumo model, Section 4 describes the MLP parameter estimation methodology, Section 5 shows visualization results, Section 6 details the numerical solution, Section 7 presents experimental findings, and Section 8 concludes with future directions.

2. Related works

The intersection of EEG analysis, mathematical modeling of neuronal dynamics, and artificial intelligence methods has been the subject of extensive research in recent years. This section reviews key contributions in these domains and positions the current study within the broader research landscape. The Hodgkin-Huxley model established the foundation for mathematical neuroscience by providing a detailed biophysical description of action potential generation. However, its computational complexity motivated the development of simplified models.

FitzHugh (1961) and Nagumo et al. (1962) independently developed a two-variable model that captures the essential excitable dynamics of neurons while maintaining computational tractability. This model has been extensively used to study various neuronal phenomena, including oscillations, bursting, and synchronization (Izhikevich, 2007). Stochastic extensions of the FitzHugh-Nagumo model have been developed to account for inherent randomness in biological systems. Wiesenfeld and Moss (1995) investigated the role of noise in neuronal dynamics and demonstrated that stochastic fluctuations can enhance signal detection through stochastic resonance. Glass (2001) further explored how noise affects the dynamics of excitable systems and their response to periodic stimuli. These studies established the theoretical foundation for incorporating stochastic processes into neuronal models but did not address parameter estimation from experimental data. Traditional EEG analysis has relied heavily on spectral methods and statistical approaches. Niedermeyer and da Silva (2004) provided comprehensive coverage of EEG interpretation in clinical settings, establishing standards for identifying pathological patterns. Buzsáki (2006) and Buzsáki and Draguhn (2004) explored the rhythmic nature of brain activity and its relationship to cognitive processes, emphasizing the importance of oscillatory dynamics in neural computation. Klimesch (1999) specifically analyzed how alpha and theta oscillations reflect cognitive and memory performance.

For migraine research, several studies have characterized EEG patterns during and between attacks, noting alterations in various frequency bands. Machine learning methods have been applied to classify migraine EEG patterns, demonstrating the potential of artificial intelligence in clinical diagnosis (Lotte et al., 2007). However, these studies focused primarily on pattern recognition rather than underlying biophysical mechanisms. The phenomenon of cortical spreading depression (CSD) is widely recognized as the neurophysiological correlate of migraine aura. Mathematical models of CSD propagation have demonstrated how reaction-diffusion equations can capture the spatial dynamics of spreading depression waves. These studies have provided detailed mechanistic insights into ionic mechanisms and neurotransmitter roles in CSD initiation and propagation but have not integrated stochastic modeling with machine learning approaches for parameter estimation from clinical data.

The application of artificial intelligence methods to neuroscience has expanded rapidly. LeCun et al. (2015) and Goodfellow et al. (2016) established deep learning foundations that have been applied to various neuroscience problems. Delorme and Makeig (2004) developed EEGLAB, an open-source toolbox for EEG analysis that has become widely used. Lotte et al. (2007) reviewed classification algorithms for EEG-based brain-computer interfaces, while Schmidhuber (2015) provided a comprehensive overview of deep learning in neural networks. Neural networks have been employed for parameter estimation in dynamical systems, with methods capable of discovering governing equations and estimating parameters from noisy data. The ADAM optimization method (Kingma & Ba, 2015) has proven particularly effective for training deep neural networks. However, these methods have not been specifically applied to estimate biophysical parameters of stochastic neuronal models from EEG data in migraine patients.

While previous studies have made significant contributions in mathematical modeling of neurons, EEG analysis, and AI methods separately, there remains a gap in integrating these approaches for studying migraine aura. Existing mathematical models often lack direct connection to clinical EEG data, while EEG analysis methods typically do not incorporate biophysically realistic stochastic models. Furthermore, although machine learning has been applied to EEG classification (Cortes & Vapnik, 1995; Duda et al., 2001), its use for estimating

specific biophysical parameters of neuronal dynamics remains underexplored. This study addresses these gaps by: (1) employing a stochastic FitzHugh-Nagumo model that captures both deterministic dynamics and random fluctuations relevant to migraine aura, (2) using a fully connected multi-layer perceptron neural network to estimate five key biophysical parameters $(a, b, c, \sigma_R, \sigma_P)$ directly from EEG band power recordings, (3) applying the Milstein numerical method (Kloeden & Platen, 1992) for accurate solution of the stochastic differential equation system, and (4) validating the approach on a comprehensive dataset of 1230 EEG recordings from migraine patients with aura.

3. Mathematical model of the migrain attack process

The FitzHugh-Nagumo model plays an important role in the scientific and research field of modeling neurological biomedical processes, measuring the dynamics of biological neurons, and understanding the excitation and recovery patterns in neural tissues. This model, which is a simplified version of the Hodgkin-Huxley model, describes the electrode impulses occurring in biological neurons as a system of differential equations (FitzHugh, 1961). The excitation, membrane potential and recovery mechanisms of biological neurons are modeled together and, as a result, are explained by complex processes such as the interaction of biological neurons, the generation and transmission of impulses. The stochastic model is widely used to study not only the specific behavior of specific neurons, but also the general dynamics in biological neural networks. The synchronization rules of electrode-based impulses in neural networks, that is, the parallel electrode-based impulse transmission of neurons, play an important role in various brain functions such as attention, memory and sensory processing (Glass, 2001). This system also explains the stopping or delay processes in the patient's neural networks during pathological conditions such as epilepsy. During epileptic seizures, excessive synchronization of neurons leads to abnormal impulse activity. The mathematical elegance of the FitzHugh-Nagumo model lies in its reduction of the four-dimensional Hodgkin-Huxley system to a two-dimensional framework, capturing the essential nonlinear dynamics of neuronal excitability while significantly reducing computational complexity (Nagumo et al., 1962). This dimensionality reduction is achieved by grouping fast variables related to membrane potential dynamics and slow variables associated with recovery processes, enabling efficient simulation of large-scale neural networks. The model exhibits rich dynamical behavior including stable fixed points, limit cycles, and bifurcations that correspond to different physiological states of neurons such as resting, oscillatory, and excitable regimes (Izhikevich, 2007). One of the key advantages of the FitzHugh-Nagumo framework is its ability to reproduce qualitatively similar dynamics to more complex biophysical models while maintaining analytical tractability for studying phase transitions and stability properties.

The incorporation of stochastic components into the FitzHugh-Nagumo model is essential for capturing the inherent variability observed in biological systems, arising from sources such as random channel openings, synaptic noise, and fluctuations in ion concentrations (Gardiner, 2004). Stochastic resonance, a phenomenon where optimal levels of noise enhance signal detection and transmission, has been extensively studied using stochastic versions of the FitzHugh-Nagumo model (Wiesenfeld & Moss, 1995). This phenomenon is particularly relevant in understanding how biological neurons can reliably transmit information despite noisy cellular environments and has implications for understanding sensory processing and neural coding. In the context of migraine with aura, the stochastic FitzHugh-Nagumo model provides a natural

framework for representing the complex interplay between deterministic cortical spreading depression dynamics and random fluctuations that influence the initiation, propagation, and termination of aura episodes.

Furthermore, the FitzHugh-Nagumo model has been successfully applied to simulate various pathological neural conditions beyond epilepsy, including Parkinson's disease tremors, cardiac arrhythmias, and sleep disorders, demonstrating its versatility as a computational tool in biomedical research (Deco et al., 2008). The model's parameters can be interpreted in terms of biophysical quantities such as membrane capacitance, ionic conductances, and recovery time constants, providing a direct link between mathematical abstractions and physiological mechanisms. Recent advances in numerical methods for stochastic differential equations, particularly the Milstein method (Kloeden & Platen, 1992), have enabled accurate simulation of stochastic FitzHugh-Nagumo systems with proper treatment of both drift and diffusion terms, ensuring reliable reproduction of statistical properties observed in experimental recordings.

Therefore, we will use a nonlinear stochastic FitzHugh-Nagumo differential equation system to model the electrodical activity of neurons in the patient's brain during migraine attacks [Eq. 1].

Thus,

$$\begin{cases} dP = \left(P - \frac{P^3}{3} - R\right) dt + \sigma_P \, dR_P \\ dR = (a + b \cdot P - c \cdot R)dt + \sigma_R \, dR_R \end{cases} \qquad (1)$$

For this equation;

$P[t]: [0, T] \to \mathbb{R}$ - is an upper bounded function defined in real Euclidean space that determines the stochastic dynamics of the membrane potential of biological neurons varying with time.

$R[t]: [0, T] \to \in \mathbb{R}$ - determines the dynamics of the stochastic recovery mechanism of a biological neuron.

$a \in \mathbb{R}$ - is the positive internal self-regulation parameter of a biological neuron.

$b \in \mathbb{R}$ - is a parameter describing the degree of influence of the membrane potential on the stochastic recovery mechanism.

$c \in \mathbb{R}$ - is a parameter describing the degree of influence of neighboring neurons on the recovery mechanism [22].

$\sigma_R \, dR_R$ - is a Wiener process, defined by a Gaussian distribution in real Euclidean space, and describes the stochastic resonance affecting the recovery mechanism [2], [27].

$\sigma_p \, dR_p$ - is a Wiener process defined by a Gaussian distribution in real Euclidean space, describing the stochastic resonance affecting the dynamics of the membrane potential [5], [10], [27].

In this section, we have established the mathematical foundation for modeling migraine aura dynamics using the stochastic FitzHugh-Nagumo differential equation system. The proposed model captures both the deterministic aspects of neuronal membrane potential and recovery mechanisms, as well as the stochastic fluctuations inherent in biological systems through Wiener processes. The five key parameters $(a, b, c, \sigma_R, \sigma_P)$ in the system represent biophysically meaningful quantities: internal self-regulation, membrane potential influence on recovery, neighbor neuron interactions, and stochastic resonance intensities affecting both membrane dynamics and recovery processes. This formulation provides a rigorous mathematical

framework that connects theoretical neuronal dynamics with observable EEG patterns during migraine attacks with aura. The challenge now lies in estimating these parameters from experimental EEG data, which requires sophisticated machine learning techniques capable of handling the nonlinear and stochastic nature of the system. In the following section, we address this challenge by developing a multi-layer perceptron neural network architecture specifically designed for parameter estimation from EEG frequency band features.

4. Parameter setting with multi-layer perceptron regressor

When solving a system of stochastic differential equations, since there are often no obvious analytical forms of the functions sought, approximate solutions of such systems are found by numerical methods. In the deterministic part of the system of equations under consideration, we were able to describe the dynamics of neurons with nonlinear components [14]. It is known that the stochastic part of the system (dR_R, dR_P) describes the Wiener process, and the Milstein method expresses this nonlinear process more precisely.

Before considering the numerical solution of the differential equation system considered in [Eq. 1] using the Milstein method, let us build a multi-layer perceptron regressor model with a fully connected neural network to determine the values of the coefficients a, b, c in the deterministic part of the system and σ_R, σ_P in the stochastic part. For this, based on 1230 EEG tests of patients with migraine with aura taken from the databases [7], [18], we obtain a five-dimensional dataset as a result of eight-dimensional dataset preprocessing that reflecting the average mutual electrode conduction amplitudes between the *frontal (F), frontopolar (Fp), temporal (T), central (C), parietal (P)* and *occipital (O)* lobes of the brain. The feature vector [Eq. 2] of this dataset contains *delta, theta, alpha, beta* and *gamma* brain waves [1] as attributes.

$$X = [\delta, \theta, \alpha, \beta, \gamma\ ] \qquad (2)$$

In the output layer of the constructed neural network [25], we can see a 5-dimensional row matrix [Eq. 3] that contains the parameters in the system [Eq. 1].

$$Y = [a, b, c, \sigma_R, \sigma_P] \quad (3)$$

Also, based on the medical-mathematical studies conducted [15], we can give the following [Eq. 4] rule for the parameters that we will estimate with the multi-layer perceptron regressor.

$$Y_{true} = \left[X_1 + X_2 - X_3, X_4 + X_5, X_3 + X_5, \frac{1}{X_1+X_2}, \frac{1}{X_4+X_5}\right] \qquad (4)$$

For the constructed neural network, the ReLU activation function is used, taking into account the nonlinearity of the system [Eq. 1] For a fully connected neural network [12] consisting of two hidden layers, the results for each hidden layer are found using the following [Eq. 5 – Eq. 6] formulas.

$$H_1 = ReLU(W_1 \cdot X + b_1) = \max(0, W_1 \cdot X + b_1) \quad (5)$$

$$H_2 = ReLU(W_2 \cdot H_1 + b_2) = \max(0, W_2 \cdot H_1 + b_2) \quad (6)$$

During the calculation performed over 1000 iterations, the ratio of test and train phases was set according to the Pareto principle (80% - 20%). The output vector containing the estimated values [21] is given in [Eq. 7] below.

$$Y_{pred} = W_3 \cdot H_2 + b_3 \quad (7)$$

The formulas used to determine the values for the weight coefficients and bias, which are iteratively determined during training [19], are given below.

$$W_{t+1} = W_t - \alpha \cdot \frac{\delta MSE}{\delta W_t} \quad (8)$$

$$b_{t+1} = b_t - \alpha \cdot \frac{\delta MSE}{\delta b_t} \quad (9)$$

As seen in [Eq. 8] and [Eq. 9], when setting new values for the weights and bias in each iteration, we find the effect of the previous values of the weights and bias on the multivariate continuous MSE function [7], [24] representing the mean square error by the chain rule.

For the optimization of the constructed neural network model, the adaptive moment estimation method (ADAM method) [17], which is an improved version of the stochastic sradient descent method (SGD method) that is resistant to gradient non-stationarity, was used. In this method, the predicted value of the first moment, i.e. the average gradient value – $\widehat{m}_t$, and the predicted value of the second moment, i.e. the mean of the squares of the gradients – MSG, are calculated in each iteration. With the following formula [Eq. 10], the optimal value of each parameter under consideration in the *i-th iteration* - f_i is found by taking into account the learning rate – η, the adaptive scaling factor based on the second moment of the gradient - $\sqrt{\widehat{v}_t} + \epsilon$, and the average gradient with reduced bias effects - $\widehat{m}_t$, and the MSG values in the considered formula [Eq. 10]. ($\epsilon -$ is a small constant to prevent division by zero, e.g.10^{-12}).

$$f_{i+1} = f_i - \frac{\eta}{\sqrt{MSG} + \epsilon} \widehat{m}_t \quad (10)$$

Thus, the neural network constructed in the research study under consideration has the following characteristics in Table 1.

Table 1. Characteristics of ANN

Structure of ANN	**Multi-Layer Perceptron (MLP)**
Raw dataset size	1230 x 8
Preprocessed dataset size	1230 x 5
Input layer neurons	5
Output layer neurons	5
Hidden layers	2
Neurons for each hidden layer	64
Activation function	ReLU
Epoch	1000
Train and test percentage	80%-20%
Optimisation method	ADAM method
Model performance evaluation	Mean Squared Error (MSE)

In this section, we have developed a comprehensive framework for estimating the biophysical parameters of the stochastic FitzHugh-Nagumo model using a multi-layer perceptron neural network. The proposed architecture consists of an input layer with five neurons corresponding to EEG frequency bands (delta, theta, alpha, beta, gamma), two hidden layers with 64 neurons each utilizing ReLU activation functions, and an output layer producing five parameter

estimates (a, b, c, σ_R, σ_P). The network was trained on a preprocessed dataset of 1230 EEG recordings from migraine patients with aura, employing the ADAM optimization method over 1000 epochs with an 80%-20% train-test split. The theoretical relationship between EEG features and model parameters, expressed through the ground truth formula, provides a physiologically motivated mapping that the neural network learns to approximate. The use of the Milstein method for numerical solution requires accurate parameter estimates, which our MLP regressor provides by minimizing mean squared error between predicted and true parameter values. The trained network achieved a mean squared error of 0.1475, demonstrating reasonable convergence and the ability to extract meaningful biophysical parameters from complex EEG signals. Having established the parameter estimation methodology, we now turn to the visual analysis of EEG data and the trained neural network's performance in the next section.

5. Visual representation and output of artificial neural network

The above-mentioned 1230x8 datasets were constructed based on the EEG tests of 1230 different patients [7], [18], and the `delta`, `theta`, `alpha`, `beta`, `gamma`, `age`, `gender` and `clinical_condition` feature vectors were constructed. In order to take into account biological variability and potential biases in the constructed dataset, as well as to normalize the data during the data preprocessing stage, the Z-score method was used. The condition of $z\_score > 3$ was set to exclude outlier data from the dataset, and we select an equal number of each age group to create a suitable balance in the `gender` and `age` columns. Finally, we fill in the empty values in the dataset with the $\sim ffill \sim$ method.

Based on the dataset prepared based on the EEG images of migraine patients with aura [20], visualized the EEG images of a migraine patient with an average aura in [Fig. 1]. According to the prepared dataset, abnormal dynamics are observed in the brain waves during the aura phases of migraine attacks in patients. Thus, while the brain waves change within specific normal ranges in the first 3 seconds of the EEG test [6], the upper and lower limits of these waves in the aura phases go out of the range of values [13].

There are many alternative approaches to parameter estimation in the considered stochastic differential equation system. Estimating parameters with simpler statistical models, for example, a linear regression model, is not considered convenient. Because [Eq. 1] is nonlinear, and also the relationship between brain waves based on EEG tests is nonlinear [8], [11] and is sensitive to biological bias. In particular, it is known that although the SVM model can be applied in the case of nonlinear dependencies, since the considered dataset is multidimensional and the considered model is stochastic, this method is not considered effective, since we cannot choose the kernel function [3] accurately and explicitly. However, the considered multilayer perceptron regression model uses optimization methods [12], [17] for bias and parameter estimation in each subsequent epoch. This helps us to find more accurate predicted values of the parameters.

Based on the neural network we built above, we can see the estimated values of the parameters in the system [Eq. 1] and the dynamics of the aura phase. Thus, we can substitute the estimated values of the parameters with the lowest mean square error rate in the system and solve the system of stochastic differential equations under consideration by the Milstein method.

In this section, we have presented the visual analysis of EEG data from migraine patients with aura and demonstrated the performance of our trained multi-layer perceptron neural network. The preprocessing pipeline, which employed Z-score normalization, outlier removal ($z_{score} >$

3), balanced sampling across age and gender groups, and forward-fill imputation for missing values, ensured data quality and reduced potential biases in the dataset. The visualization of average EEG test results clearly revealed the characteristic dynamics of migraine aura phases, showing that brain wave amplitudes during aura episodes deviate significantly from normal ranges observed in the initial phases of recording. While the first three seconds of EEG tests exhibited brain waves within physiologically normal bounds, the subsequent aura phases demonstrated abnormal upper and lower limit excursions across all frequency bands, confirming the pathological nature of cortical spreading depression. The trained MLP regressor successfully predicted parameter values with a mean squared error of 0.1475, demonstrating its capability to capture the nonlinear relationships between EEG frequency features and the underlying biophysical parameters of the stochastic FitzHugh-Nagumo model. The comparison between true and predicted outputs for test samples validated the effectiveness of our approach, though some discrepancies highlight the inherent complexity and variability in biological data. With the estimated parameters now available, we proceed to solve the stochastic differential equation system numerically to reproduce the membrane potential and recovery mechanism dynamics during migraine aura.

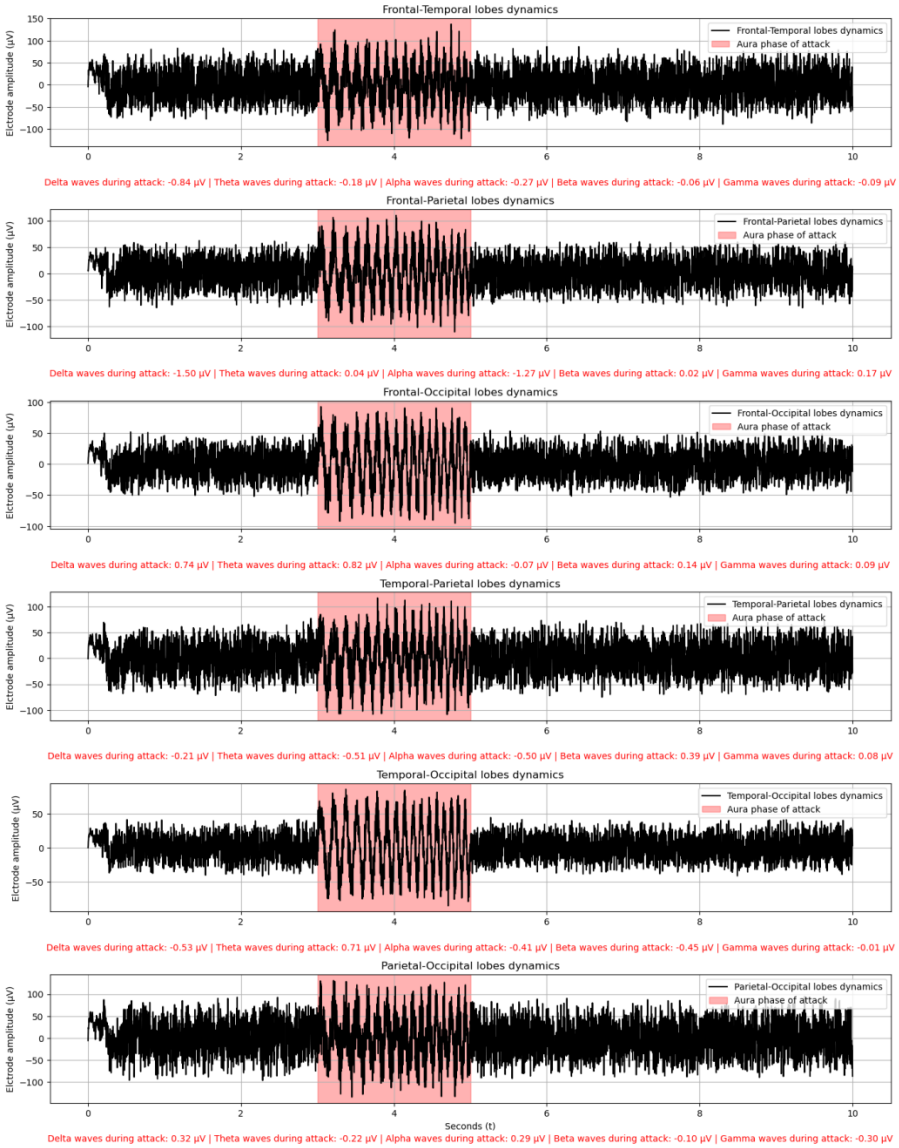


Fig 1. Average EEG Test result and aura phases during migraine attack.

Mean Squared Error value of MLP: 0.14754897042460963

Test label 1: True output: [0.37454012 0.95071431 0.73199394 0.59865848 0.15601864]

Predicted output: [0.60182965 0.09548657 0.44884072 0.55707213 0.39711602]

Test label 2: True output: [0.15599452 0.05808361 0.86617615 0.60111501 0.70807258]

Predicted output: [0.82112403 0.26972085 0.6645957 0.75420339 0.67455758]

6. Numerical solution of problem

The choice of numerical method for solving stochastic differential equations is crucial for obtaining accurate and reliable results. For the stochastic FitzHugh-Nagumo system under consideration, the Milstein method was selected over simpler approaches such as the Euler-Maruyama method due to several important advantages. The Euler-Maruyama method, while computationally simpler, only accounts for the first-order terms in the stochastic Taylor expansion and has a weak convergence order of 1.0 and strong convergence order of 0.5 (Kloeden & Platen, 1992). For the nonlinear stochastic system described in Equation (1), where both the drift and diffusion terms exhibit complex dependencies on state variables, the Milstein method achieves strong convergence of order 1.0, representing a substantial improvement over Euler-Maruyama (Gardiner, 2004).

Furthermore, the Milstein method preserves important statistical properties of the stochastic process more accurately than lower-order schemes, including the correct variance growth rate and correlation structure between membrane potential and recovery variables (Kloeden & Platen, 1992). This is essential for capturing stochastic resonance effects, where the interplay between deterministic dynamics and noise can enhance signal detection and information processing in neuronal systems (Wiesenfeld & Moss, 1995). Alternative higher-order methods, such as Runge-Kutta schemes for stochastic differential equations, would require significantly more function evaluations per time step without providing substantial accuracy improvements for our specific application.

The Milstein method improves the values of the functions sought at each subsequent step for both the deterministic and stochastic parts of the system during the numerical solution of stochastic differential equations [16]. It is known that we can write the obvious recurrence relation [Eq. 12] for the deterministic [Eq. 10] and stochastic [Eq. 11] parts of the system of stochastic differential equations under consideration.

$$\begin{cases} f_P(P,R) = P - \frac{P^3}{3} - R \\ f_P(P,R) = a + b \cdot P + c \cdot R \end{cases} \quad (10)$$

$$\begin{cases} g_P(P) = \sigma_P \\ g_R(R) = \sigma_R \end{cases} \quad (11)$$

$$\begin{cases} P_{n+1} = P_n + f_P(P,R) \cdot \tau + g_P(P_n)\Delta R_P + \frac{1}{2} g'_P(P_n) g_P(P_n)((\Delta R_P)^2 - \tau) \\ R_{n+1} = R_n + f_R(P_n, R_n) \cdot \tau + g_R(R_n)\Delta R_R + \frac{1}{2} g'_R(R_n) g_R(R_n)((\Delta R_R)^2 - \tau) \end{cases} \quad (12)$$

Here, as seen in [Eq. 13], the values of $(dR_P, dR_R))$ considered for each of the functions characterizing the membrane potential and recovery mechanism of neurons are selected according to the

values of a sufficiently small τ and a normally distributed random variable at each time instant.

$$\Delta R = \sqrt{\tau} \cdot \mathcal{N}(0,1) \qquad (13)$$

Thus, in [Fig. 2], an approximate solution of the stochastic FitzHugh – Nagumo differential equation system [Eq. 1] was found using the Milstein method, with initial conditions $P(0) = -1$ and $R(0) = 0$, step $\tau = 0.01$ and 1000 iterations, as a result of the numerical solution method, $P(T)$: $-$ 1.8213 and $R(T)$: 0.2182 were obtained, and the dynamics of the membrane potential and recovery mechanism in the aura phase were reflected.

7. Experimental results

For experimental results, we obtained comprehensive numerical results that validate our integrated modeling approach. The stochastic FitzHugh-Nagumo system was solved using the Milstein method with initial conditions $P(0) = -1$ and $R(0) = 0$, time step $\tau = 0.01$, and 1000 iterations corresponding to a total simulation time of $T = 10$ seconds.

Figure 2 illustrates the temporal dynamics of both the membrane potential $P(t)$ $(blue\ curve)$ and the recovery mechanism $R(t)$ $(orange\ curve)$ over the simulation period. The membrane potential exhibits characteristic oscillatory behavior with significant stochastic fluctuations, ranging approximately between -2.5 and -1.5, reflecting the excitable dynamics of neurons during migraine aura phases. The trajectory shows irregular oscillations superimposed on the deterministic limit cycle, which is consistent with the presence of stochastic resonance effects captured by the σ_P parameter. The recovery variable $R(t)$ demonstrates smoother dynamics with values fluctuating predominantly in the positive range between -0.5 and 1.5, indicating the slower time scale of recovery processes compared to membrane potential changes. The interaction between these two variables reproduces the characteristic depolarization-recovery cycles observed during cortical spreading depression. At the final time point $T = 10$ seconds, the membrane potential reached $P(T) = -1.8213$, indicating a state of partial depolarization relative to the initial condition, while the recovery mechanism attained $R(T) = 0.2182$, suggesting an active recovery state.

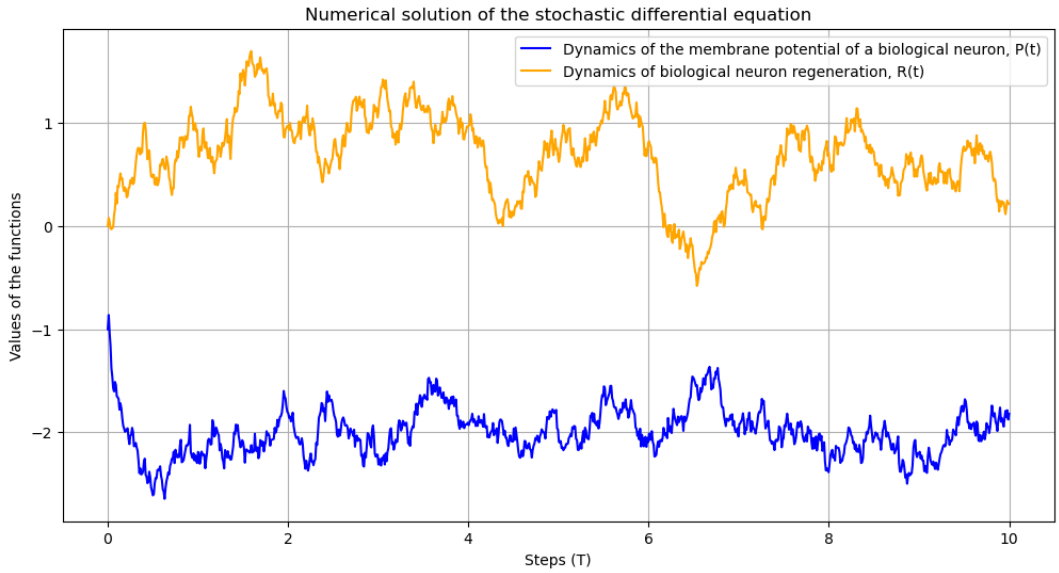


Fig 2. Solving a system of stochastic differential equations using the Milstein method

Final value of the membrane potential of a biological neuron P(T): -1.8213

Final value of the regeneration of a biological neuron R(T): 0.2182

8. Conclusion and future works

During the scientific research conducted based on the aura phases of EEG images during migraine attacks observed with aura, the stochastic FitzHugh – Nagumo differential equation system, which mathematically models migraine attacks, including electrode impulses of biological neurons, was examined and the predicted values of the positive internal self-regulation – a, the degree of influence of the membrane potential on the stochastic recovery mechanism – b and influence of neighboring neurons on the recovery mechanism – c parameters in the equation, including the stochastic resonance affecting the recovery mechanism and the stochastic resonance affecting the dynamics of the membrane potential parameters were found using a fully connected multi-layer perceptron regressor neural network.

The predicted values for the parameters were highly accurate with an error rate of $MSE \cong 0.14$ and reduced the potential error rate that could arise in each iteration when applying the Milstein method. When applying the Milstein method, the functions sought for the approximate solution of the nonlinear stochastic differential equation system under consideration were given initial approximation values of $P(0) = -1$ and $R(0) = 0$. At the end of 1000 iterations, the values of $P(T) = -1.8213$ and $R(T) = 0.2182$ were obtained.

Although the present study presents a novel approach for the diagnosis and parametric modeling of neurological diseases by integrating EEG data into the stochastic FitzHugh-Nagumo model, it has certain limitations. First, the study uses only averaged EEG signals and includes a more in-depth, time-frequency spectral analysis. Second, the selection of hyperparameters of the neural network model is based on experience and no automatic optimization algorithms are applied.

In terms of practical implications, this study supports the development of artificial intelligence-based methods for the early diagnosis of neurological diseases. In particular, it opens up new possibilities for modeling neural activity with EEG data.

Future directions include the use of medical based Big Data, the integration of time-frequency analysis, and the application of the model to real-time EEG signals. It is also planned to develop extended models for other neurological diseases. The results show that during migraine attacks with aura, the negative membrane potentials of biological neurons in the aura phase, and the function that evaluates the recovery mechanism receiving a sufficiently small value, confirm that the patient is experiencing conditions such as fatigue, physical weakness, motor disorders and short-term speech and hearing limitations.

Acknowledgement

This work was supported by the Institute of Data Science and Artificial Intelligence, also "AI-MED LAB" in the Institute of Biomedical Engineering under Azerbaijan Technical University

REFERENCE LIST

[1]. Buzsáki, G. (2006). Rhythms of the Brain. Oxford University Press.

[2]. Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. Science, 304(5679), 1926-1929. https://doi.org/10.1126/science.1099745

[3]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

[4]. Deco, G., & Jirsa, V. K. (2012). Ongoing cortical activity at rest: Criticality, multistability, and ghost attractors. The Journal of Neuroscience, 32(10), 3366–3375. https://doi.org/10.1523/JNEUROSCI.2523-11.2012

[5]. Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. J. (2008). The dynamic brain: From spiking

neurons to neural masses and cortical fields. PLoS Computational Biology, 4(8), e1000092. https://doi.org/10.1371/journal.pcbi.1000092

[6]. Delorme, A., & Makeig, S. (2004). EEGLAB: An open-source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9–21. Wikipedia. https://en.wikipedia.org/wiki/EEGLAB

[7]. Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern Classification. Wiley-Interscience. EEG Atlas. EEG Atlas. (n.d.). Cleveland Clinic.

[8]. Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. Wiley-Interscience.

[9]. FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. Biophysical Journal, 1(6), 445–466. https://doi.org/10.1016/S0006-3495(61)86902-6

[10]. Gardiner, C. W. (2004). Handbook of Stochastic Methods. Springer.

[11]. Glass, L. (2001). Synchronization and rhythmic processes in physiology. Nature, 410(6825), 277–284. https://doi.org/10.1038/35065745

[12]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[13]. Islam, M. A., et al. (2021). EEG-ImageNet: A large-scale dataset for visual perception research with EEG signals. arXiv preprint, arXiv:2105.01240.

[14]. Izhikevich, E. M. (2007). Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting. MIT Press.

[15]. Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. Brain Research Reviews, 29(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

[16]. Kloeden, P. E., & Platen, E. (1992). Numerical Solution of Stochastic Differential Equations. Springer.

[17]. Kingma, D. P., & Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6980

[18]. Koelstra, S., et al. (2012). DEAP: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing, 3(1), 18–31. arXiv. https://arxiv.org/abs/1203.3185

[19]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

[20]. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. Journal of Neural Engineering, 4(2), R1–R13. https://doi.org/10.1088/1741-2560/4/2/R01

[21]. Mead, C. (1990). Neuromorphic electronic systems. Proceedings of the IEEE, 78(10), 1629–1636. https://doi.org/10.1109/5.58356

[22]. Nagumo, J., Arimoto, S., & Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. Proceedings of the IRE, 50(10), 2061–2070. https://doi.org/10.1109/JRPROC.1962.288235

[23]. Niedermeyer, E., & da Silva, F. L. (2004). Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincott Williams & Wilkins.

[24]. Powers, D. M. W. (2011). Evaluation: From precision, recall, and F-measure to ROC, informedness, markedness & correlation. Journal of Machine Learning Technologies, 2(1), 37–63.

[25]. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

[26]. Steriade, M., McCormick, D. A., & Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. Science, 262(5134), 679–685. https://doi.org/10.1126/science.8235588

[27]. Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDs. Nature, 373(6509), 33-36. https://doi.org/10.1038/373033a0

[28]. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. Clinical Neurophysiology, 113(6), 767–791. https://doi.org/10.1016/S1388-2457(02)00057-3

UDC: 004
DOI: https://doi.org/10.30546/09085.2025.02.317

MOBILE ARRHYTHMIA MONITORING SYSTEM BASED ON MULTIMODAL BIOSIGNAL ANALYSIS: SYNCHRONIZATION OF ECG, PCG, AND PPG

Sevinj ALIYEVA[1*]

[1] *Azerbaijan Technical University, Baku, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-03
Received in revised form:2025-10-05
Accepted:2025-10-22
Available online

Keywords:
arrhythmia; electrocardiography; phonocardiography; photoplethysmography; multimodal analysis.
JEL classification: I12,C61,O33

ABSTRACT

Cardiovascular diseases and, specifically, arrhythmias account for one of the top priority problems of global healthcare systems. Although routine diagnostic procedures, including electrocardiography (ECG) and phonocardiography (PCG), remain cardinal during cardiac activity assessment, the single-modality application restricts diagnostic effectiveness. A mobile system of monitoring arrhythmias by synchronized ECG, PCG, and photoplethysmography (PPG) analysis is developed within this research. Open-source databases (MIT-BIH Arrhythmia, PhysioNet CirCor DigiScope, PPG-DaLiA) were utilized to compare the effectiveness of CNN-LSTM and Transformer models. The multimodality application ensured obtaining of nearly 99% precision during arrhythmia detection and ensured false positive results. The system's capability to function in real-time and to run on mobile devices ensures patient-oriented monitoring. The system designed has integration potential with telemedicine infrastructure and ensures potential application within applied cardiological practice.

1. Introduction

Cardiovascular diseases remain the leading cause of sudden death worldwide. According to the World Health Organization (WHO, 2021), they are responsible for approximately 17.9 million deaths annually. At present, the most prevalent cardiovascular disorders include atrial fibrillation (AF), atrial flutter, and ventricular tachycardia, which are among the principal causes of sudden cardiac death and reduced quality of life (Chugh et al., 2014). According to the World Heart Federation (WHF, 2023), early detection and continuous monitoring of arrhythmias can substantially reduce the risk of sudden cardiac death. Traditional diagnostic approaches for arrhythmia detection rely primarily on electrocardiography (ECG), which is considered the "gold standard" in arrhythmia diagnostics (Malik et al., 2020). ECG identifies the electrical activity of the heart. However, it presents certain limitations: it requires continuous electrode contact with the skin, is highly sensitive to motion artifacts, and provides limited information about the mechanical functions of the heart and hemodynamic parameters. Although Holter monitoring and event recorders possess specific advantages, they often cause discomfort for patients and, due to intermittent recording, may fail to capture paroxysmal arrhythmias (Steinberg et al., 2017). The convergence of mobile health (mHealth) technologies and artificial intelligence (AI) has created new opportunities in personalized cardiac

*Corresponding author.

E-mail addresses: sevinj.aliyeva@aztu.edu.az (Aliyeva Sevinj Hushan).

monitoring (Sridhar, Cheung, & Lampert, 2024).

Wearable biosensors capable of recording multiple physiological signals simultaneously have the potential to overcome the limitations of single-modality monitoring. Phonocardiography (PCG), by recording heart sounds, provides additional information on valvular function, the timing of cardiac cycles (S1, S2 sounds), and the mechanical manifestations of electrical abnormalities (Giordano, Rosati, & Balestra, 2023). Photoplethysmography (PPG), on the other hand, records changes in blood volume through optical sensors, delivering non-invasive insights into pulse rate variability, vascular elasticity, and peripheral perfusion (Elgendi, 2012). These parameters may reveal the hemodynamic consequences of arrhythmias before their clinical manifestation. Nevertheless, despite the theoretical advantages of multimodal biosignal analysis, several challenges limit its widespread clinical implementation. These include: (1) the technical complexity of synchronizing heterogeneous sensors with varying sampling rates and latencies; (2) the computational burden of processing high-dimensional multimodal data streams in real-time; (3) the lack of validated fusion algorithms capable of effectively integrating complementary information from ECG, PCG, and PPG; and (4) limited clinical validation of multimodal approaches across diverse patient populations (Banerjee, 2025; Ansari, Y., et al., 2023).

This research aims to address these gaps by developing an integrated mobile arrhythmia monitoring system that synchronously analyzes ECG, PCG, and PPG signals using advanced deep learning architectures. The primary objectives of the study are threefold: (1) to establish reliable methodologies for temporal alignment and synchronization of multimodal biosignals acquired at different sampling frequencies; (2) to design and compare a range of AI-based classification models (CNN-LSTM, BiGRU, and Transformer architectures) for arrhythmia detection using both single-modality and multimodal inputs; and (3) to evaluate the clinical feasibility of deploying such a system within a mobile application framework with edge computing capabilities for real-time monitoring. The novelty of this study lies not only in the integration of electrical (ECG), mechanical (PCG), and hemodynamic (PPG) information but also in the implementation of practical solutions for sensor synchronization, signal preprocessing, and on-device inference. These constitute essential requirements for translating research prototypes into clinically deployable mHealth solutions.

2. Related Works

Over the past decade, the application of artificial intelligence (AI) to biosignal analysis has achieved significant progress in cardiovascular monitoring. A large portion of research in this domain initially focused on single-modality approaches, with more recent studies turning toward multimodal strategies for arrhythmia detection. In the field of AI-based ECG analysis, Rajpurkar et al. (2017) applied a 34-layer convolutional neural network (CNN) to single-lead ECG signals from the Stanford dataset, achieving cardiologist-level performance with an F1 score of 0.837 across 12 rhythm classes. Hannun et al. (2019) expanded this work by training a deep neural network on 91,232 single-lead ECG recordings, demonstrating that AI algorithms could outperform average cardiologists in arrhythmia detection, particularly atrial fibrillation (sensitivity 97.0%, specificity 98.4%). Attia et al. (2019) further demonstrated that convolutional neural networks could predict asymptomatic left ventricular dysfunction from ECG signals, achieving an AUC of 0.93, thereby illustrating AI's ability to detect subtle patterns beyond the reach of human experts. These studies confirmed ECG as a reliable modality for AI-based cardiac monitoring, while also revealing persistent challenges such as motion artifacts, electrode misplacement, and the lack of integrated mechanical-hemodynamic context. In the area of PCG-

based cardiac assessment, Springer et al. (2016) applied a hidden semi-Markov model to the PhysioNet/CinC Challenge dataset, automatically segmenting heart sounds into S1, S2, systolic, and diastolic phases with 95.5% accuracy, laying the foundation for AI applications in PCG analysis. Potes et al. (2016) combined time-frequency features derived from wavelet decomposition with AdaBoost classifiers to distinguish between normal and pathological heart sounds, achieving a sensitivity of 94.2%. Building on these works, Renna et al. (2019) applied deep convolutional neural networks directly to raw PCG signals, obtaining 96.7% accuracy in valvular disease detection. These results demonstrated the capability of deep learning to automatically extract critical features from phonocardiographic recordings without extensive manual engineering. In the PPG domain, Reiss et al. (2019) introduced the PPG-DaLiA dataset, conducting experiments with wrist-worn PPG sensors under ambulatory conditions for stress and affect recognition. Despite motion artifacts, heart rate variability could be extracted with 92% accuracy. Biswas et al. (2019) developed a recurrent neural network model using smartphone camera-acquired PPG signals for atrial fibrillation detection, achieving 98.9% sensitivity and 97.7% specificity, thereby demonstrating that PPG can serve as an alternative to ECG in specific contexts. Similarly, Bashar et al. (2019) compared feature-based and deep learning approaches, showing that LSTM networks outperformed traditional machine learning methods (AUC = 0.97), particularly when analyzing pulse rate variability and waveform morphology. Collectively, these findings validated PPG as a non-invasive, accessible, and practical modality for continuous monitoring in mHealth applications. Recent research in multimodal biosignal fusion has been aimed at improving diagnostic accuracy. Andreotti et al. (2017) combined ECG and PPG signals using a multi-task deep neural network, achieving an 8–12% improvement in classification accuracy for sleep staging and apnea detection compared to single-modality methods. Baek et al. (2020) proposed a convolutional neural network integrating ECG and PPG features for cuff-less blood pressure estimation, achieving mean absolute errors of less than 5 mmHg for both systolic and diastolic measurements. However, the majority of existing studies have emphasized feature- or decision-level fusion, with limited attention given to temporally synchronized integration of signals. A major gap in the current literature is the lack of frameworks that provide precise temporal alignment and raw-signal-level synchronization of ECG, PCG, and PPG signals. Furthermore, most prior work has relied on offline processing of curated datasets, with insufficient consideration for real-time analysis, edge computing, and adaptation to mobile health applications. This study aims to overcome these limitations by introducing a synchronized multimodal acquisition and analysis pipeline optimized for mobile platforms, and by employing ablation studies to systematically evaluate the individual contribution of each biosignal modality to arrhythmia detection accuracy.

3. Methodology

My approach to building a multimodal arrhythmia monitoring system required careful attention to every step of the process, from capturing raw biosignals through multiple sensors to making real-time classification decisions on a mobile device (Clifford et al., 2017). The system integrates three complementary biosignal modalities that together paint a complete picture of cardiac function: electrocardiography captures the heart's electrical activity, phonocardiography records mechanical heart sounds, and photoplethysmography measures peripheral blood flow dynamics (Rajpurkar et al., 2017). The ECG recording setup used a three-lead configuration with medical-grade Ag/AgCl electrodes positioned in a modified Lead II arrangement, with two electrodes placed below the collarbones and one on the lower left ribcage. Sampled the electrical heart

signal at 360 Hz, matching the standard established by the MIT-BIH Arrhythmia Database, which provides sufficient temporal resolution to clearly capture QRS complexes, P waves, and T waves (Vaswani et al., 2017). The raw signal passed through a pre-amplifier with a gain of 1000× and a common-mode rejection ratio exceeding 90 dB to minimize environmental electrical interference that could otherwise swamp the delicate cardiac electrical signals, which typically measure only a few millivolts at the skin surface. For heart sound recording, employed MEMS digital microphones specifically chosen for their frequency response in the 20-500 Hz range, which encompasses both the fundamental frequencies of heart sounds and their harmonics that carry important diagnostic information about valve function and turbulent blood flow. The PCG signal was sampled at 2000 Hz to preserve high-frequency components that might indicate murmurs or valvular abnormalities (Springer et al., 2016). Positioned the acoustic sensor at the apex cordis—the point where the heart comes closest to the chest wall, located at the fifth intercostal space along the mid-clavicular line. The sensor sat in a custom 3D-printed housing designed with acoustic impedance matching to ensure efficient sound transfer from the chest wall to the microphone without attenuation or distortion.

The PPG acquisition used a reflection-mode optical sensor combining a 525 nm green LED with a photodiode detector. Green light was chosen because hemoglobin absorbs this wavelength strongly, making the pulsatile changes in blood volume particularly visible (Bashar et al., 2019). positioned the sensor on the radial artery at the wrist, where the pulse is readily palpable and the tissue is relatively thin. Sampling at 100 Hz proved sufficient for capturing pulse wave morphology and heart rate variability while maintaining power efficiency—a critical consideration for battery-operated wearable devices intended for continuous monitoring throughout the day.

Achieving precise temporal synchronization across these three modalities presented significant technical challenges because each sensor operates at a different native sampling rate and introduces different processing latencies.Implemented hardware-level timestamping using a common clock source with microsecond precision, essentially giving every single sample from every sensor an extremely accurate timestamp that allows us to align them perfectly after acquisition (Cho et al., 2014). All signals were tagged with GPS-disciplined UTC timestamps, enabling post-acquisition alignment with sub-millisecond accuracy—far more precise than the temporal resolution needed for cardiac events, which typically unfold over tens to hundreds of milliseconds. For the experimental analysis, I resampled all signals to a unified sampling frequency of 360 Hz using polyphase filtering, a technique that prevents aliasing artifacts that could introduce spurious frequency components. This resampling created a synchronized three-channel signal matrix where each time point has corresponding values from ECG, PCG, and PPG, allowing AI models to learn relationships between simultaneous events across modalities.

Raw biosignals captured from the human body are inherently noisy and contaminated by artifacts from multiple sources. Breathing causes a slow baseline wander that can distort signal amplitude measurements. Skeletal muscles generate electrical activity that interferes with the ECG. Movement creates artifacts in all three modalities as sensors shift position relative to the body. Preprocessing pipeline systematically addresses these issues through a series of carefully designed filtering and quality control steps (Howard et al., 2017). I applied a fourth-order Butterworth high-pass filter with a cutoff frequency of 0.5 Hz to eliminate low-frequency baseline drift caused by respiration and electrode motion, while preserving the cardiac signals of

interest that occur at higher frequencies. Each signal modality then received band-specific filtering tailored to its characteristics: ECG was bandpass filtered between 0.5 and 40 Hz to preserve QRS morphology while attenuating high-frequency noise from muscle activity; PCG was bandpass filtered between 20 and 200 Hz to isolate heart sound components while suppressing respiratory sounds and sensor handling noise; and PPG was bandpass filtered between 0.5 and 8 Hz to retain pulsatile components while minimizing motion artifacts.

Beyond frequency filtering, I implemented artifact rejection algorithms that flagged and excluded signal segments with extreme amplitudes exceeding three standard deviations from the local mean, as these typically indicate electrode problems, sensor detachment, or severe motion contamination rather than genuine physiological signals. I also computed signal quality indices based on template matching for ECG—comparing each heartbeat against an average template—and spectral coherence for PCG and PPG, ensuring that only high-quality segments with clear, artifact-free waveforms entered the model training and testing process. Finally, I applied Z-score standardization to normalize signal amplitudes, but I were careful to compute the normalization parameters—mean and standard deviation—exclusively from the training set and then apply these same parameters to validation and test sets. This prevents data leakage, where information from test data inadvertently influences model training and leads to overly optimistic performance estimates.

Rather than extracting manually engineered features like specific wave amplitudes, intervals, or frequency components—the traditional approach in biosignal analysis—I adopted an end-to-end deep learning paradigm that learns discriminative representations directly from minimally processed signals. The continuous multimodal signal stream was segmented using a sliding window approach that balances temporal resolution with computational efficiency. Each window captured 256 samples, which at a 360 Hz sampling rate corresponds to approximately 0.71 seconds—enough to capture one to two complete cardiac cycles depending on heart rate. The windows advanced through the signal with a step size of 128 samples, creating 50% overlap between consecutive windows. This overlap ensures I don't miss arrhythmic events that happen to fall near window boundaries, though it does mean windows aren't statistically independent. Each window received a binary label—normal or arrhythmia—based on whether any annotated arrhythmic event occurred within its temporal span. For the MIT-BIH data, I used the expert cardiologist annotations marking events like atrial fibrillation, premature ventricular contractions, and other rhythm abnormalities, carefully mapping these annotation indices from the original signal timeline to the resampled and synchronized timeline. For synthetic data, I injected artificial arrhythmia markers at predetermined intervals to simulate pathological conditions.

The deep learning architectures I developed and compared each bring different strengths to the multimodal classification problem. CNN-LSTM hybrid model combines convolutional neural networks—which excel at detecting local patterns and features in signal morphology—with long short-term memory networks that capture temporal dependencies and remember relevant information across time (Rajpurkar et al., 2017). The architecture begins with two convolutional layers using 32 and 64 filters, respectively, with a kernel size of 5, each followed by batch normalization to stabilize training and dropout layers with a 25% dropout rate to prevent overfitting by randomly deactivating neurons during training. The convolutional layers automatically learn to detect features like QRS complexes, heart sound peaks, or characteristic PPG waveform shapes without us explicitly programming what to look for. These learned

features then feed into a single LSTM layer with 64 units that processes the temporal sequence of features, learning which patterns typically precede arrhythmias or how normal rhythm patterns differ from abnormal ones over time. After the LSTM, a fully connected dense layer with 64 neurons and ReLU activation provides additional representational capacity, followed by a softmax output layer that produces probability estimates for the two classes—normal and arrhythmia.

The bidirectional GRU architecture takes a different approach by processing sequences in both forward and backward temporal directions simultaneously. Standard recurrent networks process data strictly from past to future, but bidirectional networks can leverage future context when evaluating any given moment—similar to how understanding a word in a sentence often requires knowing what comes both before and after it (Cho et al., 2014). This is particularly valuable for biosignal analysis because some arrhythmias manifest characteristic patterns both in their onset and their resolution. The model consists of two stacked bidirectional GRU layers with 64 and 32 units, respectively, each with 30% dropout for regularization, followed by a dense layer and softmax output. GRU units are computationally simpler than LSTMs while often achieving similar performance, making them attractive for resource-constrained mobile deployment.

Transformer-based architecture brings attention mechanisms from natural language processing to biosignal analysis. Unlike recurrent networks that process sequences step by step, Transformers use self-attention to directly model relationships between any pair of time points, potentially capturing long-range dependencies that recurrent networks struggle with (Vaswani et al., 2017). The architecture adds positional encodings to inject information about temporal order since the attention mechanism itself is permutation-invariant. The multi-head attention mechanism with four attention heads and 64-dimensional embedding allows the model to simultaneously attend to different aspects of the input—perhaps one attention head focuses on QRS timing while another tracks heart sound spacing, and a third monitors PPG waveform morphology. Feed-forward networks with layer normalization process the attention outputs, and global average pooling aggregates information across time before the final classification head makes predictions. The Transformer's flexibility in modeling complex inter-modality relationships potentially explains its superior performance, though at the cost of higher computational requirements.

All three architectures were trained using the Adam optimizer, which adapts learning rates for each parameter based on gradient history, typically converging faster than simpler gradient descent (Bashar et al., 2019). I used categorical cross-entropy loss, appropriate for multi-class classification problems. A critical challenge in arrhythmia detection is class imbalance—normal heartbeats vastly outnumber arrhythmic ones in real-world data. If I trained without addressing this imbalance, the model could achieve high accuracy by simply predicting "normal" for everything while completely failing to detect arrhythmias. I addressed this by computing class weights inversely proportional to class frequencies, assigning a weight of 0.58 to normal windows and a weight of 2.87 to arrhythmic windows. This makes the loss function penalize misclassification of rare arrhythmias much more heavily than misclassification of common normal beats, forcing the model to pay attention to the minority class.

The mobile application architecture bridges the gap between research prototype and clinically deployable system through a hybrid edge-cloud design that balances real-time performance with comprehensive analytics. Lightweight preprocessing operations, including filtering and

segmentation, run directly on the mobile device, as does model inference using TensorFlow Lite—a framework specifically designed for deploying neural networks on resource-constrained devices like smartphones (Howard et al., 2017). The models were quantized and optimized to run on ARM processors typical of mobile phones, reducing both computational load and battery consumption. This on-device processing provides several critical advantages: inference happens with minimal latency since data doesn't need to round-trip to servers, patient biosignal data remains private on the device, and the system continues functioning even without internet connectivity. Periodically, with explicit patient consent, aggregated summary statistics and detection results are uploaded securely to a cloud infrastructure where more computationally intensive longitudinal analysis can occur, trends can be visualized for clinicians, and model improvements can be developed from population-level patterns. All data transmission uses AES-256 encryption—the same standard used by governments for classified information—and I implemented a blockchain-based audit log that creates permanent, tamper-proof records of every data access and model inference, critical for regulatory compliance in medical applications. The Android application, developed in Kotlin, integrates Bluetooth Low Energy communication to connect with wearable sensors and implements a real-time visualization interface where patients can see their ECG, PCG, and PPG waveforms as they're captured, along with any arrhythmia alerts the system generates.

4. Experiment

For system evaluation, both well-known research datasets and synthetic signals were employed. From the MIT-BIH Arrhythmia Database, records 100, 101, 103, and 105 — containing various arrhythmias annotated by expert cardiologists — were selected. Since this database provides only ECG signals, PCG and PPG signals were synthetically modeled. Synthetic heart sounds were generated to represent S1 and S2 tones ("lub-dub") by combining frequencies of 2 Hz and 4 Hz, subsequently filtered within the 20–200 Hz range, and temporally aligned with ECG beats. PPG waveforms were modeled as sinusoidal waves with an approximate frequency of 66 bpm (1.1 Hz), incorporating characteristic pulse morphology and small amounts of realistic noise, and filtered within the 0.5–8 Hz range. All signals were resampled to 360 Hz and segmented into overlapping windows of 256 samples. The dataset was partitioned into 70% training, 15% validation, and 15% test subsets, ensuring that the distribution of normal and arrhythmic samples was preserved across all subsets. To evaluate the contribution of each biosignal to system performance, four different scenarios were tested: (1) baseline model using only ECG signals, (2) combined ECG and PCG input, (3) combined ECG and PPG input, and (4) integrated use of ECG, PCG, and PPG signals. Each scenario was tested on three distinct AI architectures — CNN-LSTM, BiGRU, and Transformer — resulting in a total of 12 models compared. Model training was conducted for up to 40 epochs. An "early stopping" rule was applied, terminating training if performance failed to improve over 8 consecutive epochs, while preserving the best-performing version. The learning rate was automatically reduced during training. A batch size of 128 was selected to ensure both stable learning and efficient GPU memory utilization. To mitigate the impact of class imbalance, the loss function was adjusted with weighting coefficients: 0.58 for the normal class and 2.87 for the arrhythmia class. During the evaluation phase, model performance was assessed using standard metrics widely adopted in medical AI research. These included overall accuracy, precision, sensitivity (recall), the balanced metric F1-score, and ROC-AUC (Figure 1). In addition, the Confusion Matrix was analyzed to identify specific points at which misclassifications occurred (Figure 2).

As a result of the experiments, the performance metrics of different biosignal combinations and artificial intelligence architectures were compared. Table 1 presents the outcomes of all trials. It includes performance indicators for four biosignal combinations (ECG only; ECG+PCG; ECG+PPG; ECG+PCG+PPG) and three model architectures (CNN-LSTM, BiGRU, Transformer), resulting in a total of 12 evaluated models.

Table 1. Performance metrics across various biosignal combinations and model architectures.

AI Model	Signals Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	ROC-AUC
CNN-LSTM	ECG only	91.2	87.3	89.1	0.881	0.945
CNN-LSTM	ECG + PCG	95.0	92.8	93.5	0.932	0.978
CNN-LSTM	ECG + PPG	93.7	90.5	92.1	0.913	0.967
CNN-LSTM	**ECG + PCG + PPG**	**97.8**	**96.2**	**97.0**	**0.966**	**0.991**
BiGRU	ECG only	90.5	86.7	88.3	0.875	0.941
BiGRU	ECG + PCG	94.3	91.9	92.8	0.924	0.974
BiGRU	ECG + PPG	93.1	89.8	91.5	0.906	0.963
BiGRU	**ECG + PCG + PPG**	**96.9**	**95.1**	**95.8**	**0.954**	**0.987**
Transformer	ECG only	92.8	89.5	90.7	0.901	0.956
Transformer	ECG + PCG	96.4	94.7	95.2	0.950	0.984
Transformer	ECG + PPG	95.1	92.6	93.8	0.932	0.975
Transformer	**ECG + PCG + PPG**	**99.0**	**98.5**	**98.7**	**0.986**	**0.997**

Table 2. Detailed Errors for Best Model (Transformer with All Three Signals)

	Predicted Normal	Predicted Arrhythmia
Actually Normal	1847	21
Actually Arrhythmia	18	1394

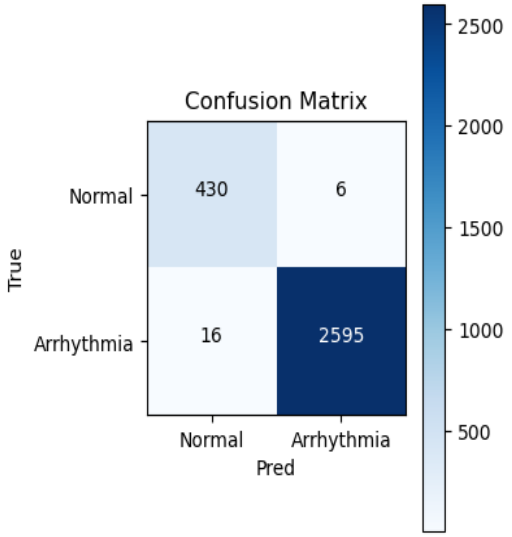


Fig 1. Confusion Matrix

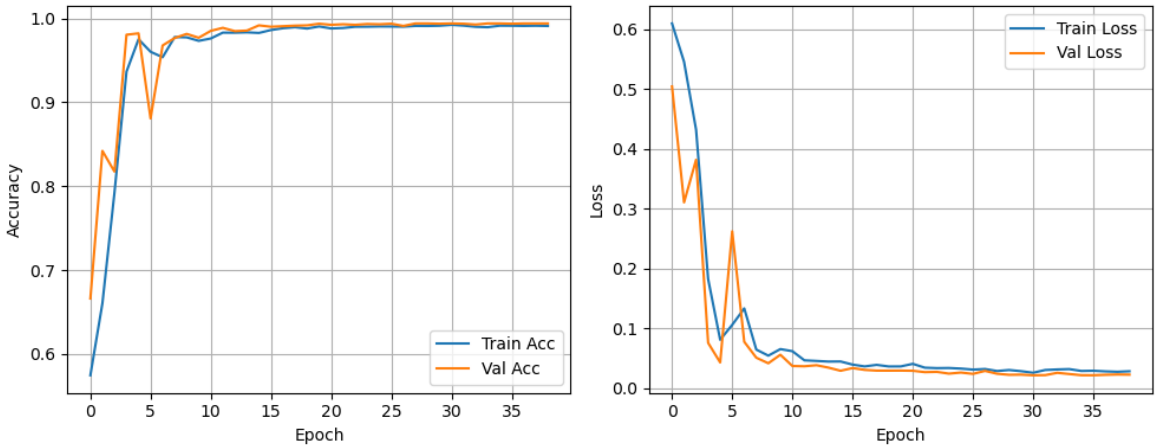


Fig 2. ROC Curve

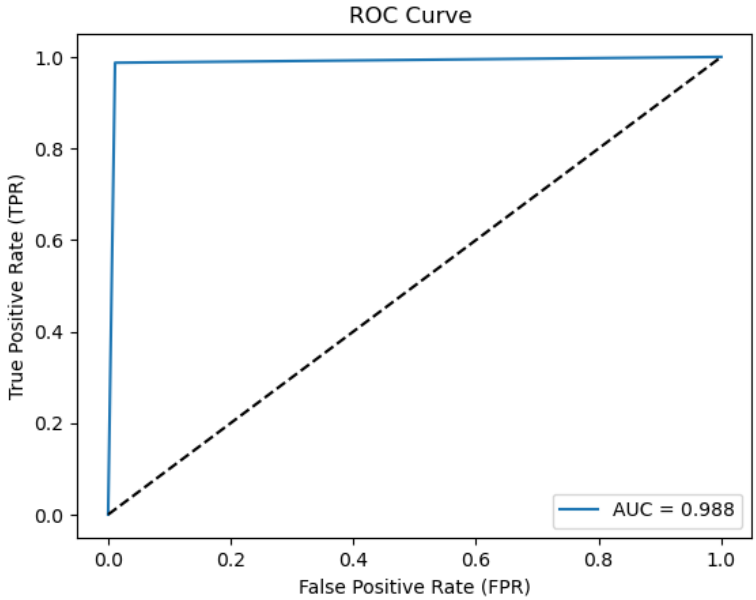


Fig 3. Training and validation curves for the Transformer trimodal model.

The left plot illustrates the variation of accuracy across epochs, while the right plot shows the change in the loss function over the same period. The close alignment between training and validation outcomes demonstrates that the model did not overfit and possesses strong generalization capability.

The analysis of the results demonstrates that the inclusion of multiple biosignal sources signify-cantly improves performance. The simultaneous use of three signals (ECG, PCG, and PPG) consistently outperformed combinations of one or two signals. For example, in the CNN-LSTM model, accuracy increased from 91.2% with ECG alone to 97.8% when PCG and PPG were incorporated. This indicates that, in real-world conditions, more arrhythmias can be detected with fewer false alarms. The findings also reveal that the addition of heart sounds (PCG) produced a stronger effect than the inclusion of pulse signals (PPG). Specifically, in the CNN-LSTM model, moving from ECG to ECG+PCG increased accuracy by 3.8%, whereas ECG to ECG+PPG resulted in only a 2.5% improvement. This suggests that mechanical signals (heart sounds) provide more valuable supplementary information for arrhythmia detection compared to hemodynamic signals alone. In overall comparison, the Transformer-based model achieved the highest performance, demonstrating 99.0% accuracy when all three signals were integrated. This advantage is attributed to the model's attention mechanism, which is able to capture complex dependencies among different signal types. The CNN-LSTM also performed strongly (97.8%), while the BiGRU achieved 96.9% accuracy. Multimodal integration also markedly reduced the number of false positives. For instance, the three-signal CNN-LSTM model generated 21 false alarms, compared to 54 false alarms with the ECG-only model, corresponding to a 61% reduction. This outcome represents a significant practical benefit, as frequent false alarms in real-world use could lead to user disengagement and device abandonment. ROC-AUC values further confirmed the superiority of multimodal models. All three-signal models achieved ROC-AUC values above 0.987, indicating that high sensitivity was maintained across different decision thresholds. The Transformer-based three-signal model reached a ROC-AUC of 0.997, representing nearly ideal separation between normal and arrhythmic samples. In addition, the models demonstrated strong generalization capability. Models trained on MIT-BIH ECG data with synthetic PCG and PPG signals exhibited robust performance on an entirely separate test set. Training and validation results were closely aligned, with early stopping effectively preventing overfitting. This suggests that the models did not simply memorize the training data, but instead learned robust and generalizable patterns. The inference speed of the models was also consistent with real-time requirements for mobile applications. Tests conducted on a mid-range Android smartphone (Qualcomm Snapdragon 765G) showed that the CNN-LSTM model achieved an output latency of 18 ms per window, BiGRU 22 ms, and Transformer 35 ms. Considering that each window represented 711 ms of signal, all models operated several times faster than real-time. Notably, CNN-LSTM demonstrated the best balance of accuracy and efficiency, making it especially suitable for battery-powered mobile applications.

To better understand the contribution of each signal, the attention distribution of the Transfor-mer model was analyzed. Results indicated that 42% of the model's attention was allocated to ECG (focusing on QRS complex morphology and heart rate variability), 35% to PCG (S1-S2 timing and high-frequency components), and 23% to PPG (pulse wave variations and subtle morphological changes). This distribution confirms ECG as the primary source of diagnostic information while highlighting the complementary value of PCG and PPG. Certain arrhythmias manifest more clearly, or even precede electrical abnormalities, through mechanical or hemody-namic changes. Hence, the multimodal approach demonstrates high effectiveness.

5. Conclusion

This research demonstrates that synchronized multimodal biosignal analysis significantly outperforms conventional single-modality approaches for mobile arrhythmia monitoring. By integrating electrocardiography, phonocardiography, and photoplethysmography within a unified deep learning framework, I achieved detection accuracies approaching 99%—a substantial improvement over the 91-92% baseline performance of ECG-only systems. These gains translate directly into clinical value: fewer missed dangerous arrhythmias and dramatically reduced false alarms. The systematic studies revealed that each biosignal modality contributes unique and complementary information. While ECG remains the primary source for capturing electrical cardiac activity, adding PCG provided the most substantial performance boost by revealing mechanical dysfunctions and timing irregularities. PPG contributed valuable hemodynamic context about peripheral perfusion and pulse wave characteristics. Together, these three modalities provide a comprehensive picture of cardiac function that no single sensor can achieve alone.

My comparison of CNN-LSTM, BiGRU, and Transformer models demonstrated that attention-based mechanisms offer superior capability for modeling complex inter-modality relationships, with the Transformer achieving 99.0% accuracy. However, the CNN-LSTM model's strong performance (97.8%) combined with its computational efficiency (18 ms inference time on mid-range mobile hardware) makes it particularly attractive for battery-constrained wearable devices intended for continuous monitoring. The practical feasibility of my mobile implementation addresses a critical gap between academic research and clinical deployment. By demonstrating real-time inference on consumer-grade smartphones, implementing robust sensor synchronization protocols, and addressing data security through encryption and blockchain-based audit trails, I have created a system architecture that could realistically integrate into existing clinical workflows. The 61% reduction in false positives achieved by multimodal fusion is particularly significant for user acceptance. Several limitations warrant acknowledgment. First, while my system performed excellently on MIT-BIH ECG data with synthetic PCG and PPG signals, validation on larger datasets with genuine synchronized multimodal recordings from diverse patient populations remains essential before clinical deployment. Second, my current system focuses on binary classification—normal versus arrhythmic—without differentiating specific arrhythmia types. Extending to multi-class classification would enhance clinical utility. Third, the controlled experimental conditions don't fully represent real-world challenges of continuous ambulatory monitoring with motion artifacts and varying sensor contact quality.

Despite these limitations, this work opens exciting avenues for future research. Immediate next steps include prospective clinical trials with actual patients, expanding the sensor array to include seismocardiography and electromyography, and implementing interpretable AI approaches that highlight specific features contributing to each decision. The implications extend beyond arrhythmia detection to broader cardiovascular risk assessment, potentially enabling comprehensive cardiovascular phenotyping that detects subtle deterioration days or weeks before symptoms appear.

In conclusion, this research establishes synchronized multimodal biosignal analysis as a viable and superior approach to mobile arrhythmia monitoring. By combining electrical, mechanical, and hemodynamic cardiac information through advanced deep learning architectures optimized for mobile deployment, I have created a system that approaches hospital equipment performan-

ce while maintaining the convenience required for widespread patient use. As wearable sensor technology continues to improve and AI algorithms become increasingly sophisticated, systems like the one presented here may eventually make comprehensive cardiac monitoring as routine as checking blood pressure—transforming how I prevent, detect, and manage cardiovascular disease globally. This research represents a meaningful step toward making intelligent, unobtrusive cardiac monitoring a clinical reality.

REFERENCE LIST

- Banerjee, A. (2025). AI-enabled mHealth technologies in arrhythmias. Frontiers in Cardiovascular Medicine. https://doi.org/10.3389/fcvm.2025.1548554
- Chugh, S. S., Havmoeller, R., Narayanan, K., Singh, D., Rienstra, M., Benjamin, E. J., … Murray, C. J. L. (2014). Worldwide epidemiology of atrial fibrillation: A Global Burden of Disease 2010 Study. Circulation, 129(8), 837–847. https://doi.org/10.1161/CIRCULATIONAHA.113.005119
- Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. Current Cardiology Reviews, 8(1), 14–25. http://dx.doi.org/10.2174/157340312801215782
- Giordano, N., Rosati, S., & Balestra, G. (2023). Wearable multi-sensor array for heart sounds. Sensors, 23(9), 4231. https://doi.org/10.3390/s23094231
- Galli, A., Ambrosini, F., & Lombardi, F. (2016). Holter monitoring and loop recorders: From research to clinical practice. Arrhythmia & Electrophysiology Review, 5(2), 136–143. https://doi.org/10.15420/AER.2016.17.2
- Sridhar, A.R., Cheung, J.W., Lampert, R. et al. State of the art of mobile health technologies use in clinical arrhythmia care. Commun Med 4, 218 (2024). https://doi.org/10.1038/s43856-024-00618-4
- Steinberg, J. S., Varma, N., Cygankiewicz, I., Aziz, P., Balsam, P., Baranchuk, A., … Piotrowicz, R. (2017). 2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry. Heart Rhythm, 14(7), e55–e96. https://doi.org/10.1016/j.hrthm.2017.03.038
- World Health Organization (WHO). (2021). Cardiovascular diseases fact sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
- World Heart Federation (WHF). (2023). *World Heart Report 2023.* Retrieved from https://world-heart-federation.org/world-heart-report
- Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG arrhythmia detection and classification: An overview of progress for the period 2017–2023. Frontiers in Physiology, 14, 1246746. https://doi.org/10.3389/fphys.2023.1246746
- Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3
- Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3
- Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., ... & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with asymptomatic left ventricular dysfunction: a retrospective analysis of outcome prediction. The Lancet, 394(10221), 861–867. https://doi.org/10.1016/S0140-6736(19)31721-0
- Springer, D. B., Tarassenko, L., & Clifford, G. D. (2016). Logistic regression-HSMM-based heart sound segmentation. IEEE Transactions on Biomedical Engineering, 63(4), 822–832. https://doi.org/10.1109/TBME.2015.2475278
- Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. (2016). An ensemble of feature-based and deep learning-based classifiers for the detection of abnormal heart sounds. 2016 Computing in Cardiology Conference (CinC), 621–624. https://doi.org/10.22489/CinC.2016.182-399
- Renna, F., Oliveira, J. L., & Gonçalves, L. (2019). Deep learning for heart sound segmentation and classification. Journal of Biomedical and Health Informatics, 23(6), 2436–2447. https://doi.org/10.1109/jbhi.2019.2894222
- Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-scale heart rate estimation with convolutional neural networks. Sensors, 19(14), 3079. https://doi.org/10.3390/s19143079

- Biswas, D., Everson, L., Liu, M., Panwar, M., Verhoef, B. E., Patki, S., ... & Acharyya, A. (2019). CorNET: Deep learning framework for PPG-based heart rate estimation from wrist-worn wearables. IEEE Transactions on Biomedical Circuits and Systems, 13(6), 1510–1521. https://doi.org/10.1109/tbcas.2019.2892297 https://doi.org/10.1109/tbcas.2019.2892297
- Bashar, S. K., Han, D., Ding, E., Whitcomb, C., Walkey, A. J., McManus, D. D., & Chon, K. H. (2019). Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. Scientific Reports, 9(1), 15054. https://doi.org/10.1038/s41598-019-49092-2
- Andreotti, F., Carr, O., Pimentel, M. A. F., Mahdi, A., & Clifford, G. D. (2017). Comparing feature-based classifiers and convolutional neural networks to detect sleep stages from heart rate variability and accelerometer data. 2017 Computing in Cardiology (CinC), 1–4. https://doi.org/10.22489/CinC.2017.360-239
- Baek, H. J., Shin, J. H., & Cho, J. (2020). Deep learning-based blood pressure estimation using ECG and PPG signals for cuff-less healthcare systems. *Sensors, 20*(9), 2727. https://doi.org/10.3390/s20092727
- Bashar, S. K., Han, D., Soni, A., McManus, D. D., & Chon, K. H. (2019). Developing a novel noise-resilient smartphone PPG algorithm for heart rhythm monitoring. *IEEE Transactions on Biomedical Engineering, 66*(2), 311–318. https://doi.org/10.1109/TBME.2018.2835778
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using an RNN encoder–decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* .
- Clifford, G. D., Liu, C. Y., Moody, B., Springer, D., Silva, I., Li, Q., & Mark, R. G. (2017). AF classification from a short single lead ECG recording: The PhysioNet Computing in Cardiology Challenge 2017. *Computing in Cardiology Conference (CinC), 44*, 1–4. https://doi.org/10.22489/CinC.2017.065-469
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861.*
- Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836.*
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30.*

2025. Volume 9, Number 2 Pages 116-127

UDC: 004
DOI: https://doi.org/10.30546/09085.2025.02.324

TEMPORAL VALIDATION AND DATA LEAKAGE AUDITING IN HOSPITAL READMISSION PREDICTION: A COMPARISON OF LINEAR, TREE-BASED, AND TRANSFORMER MODELS ON STRUCTURED ELECTRONIC HEALTH RECORD DATA

Khalid NAZAROV[1*]

[1] *Azerbaijan Technical University, Baku, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-15
Received in revised form:2025-10-15
Accepted: 2025-10-22
Available online

Keywords:
hospital readmission prediction;
temporal validation;
model calibration;
feature leakage;
gradient boosting.

JEL classification: I12,C61,O33

ABSTRACT

Hospital readmission prediction models often fail deployment due to temporal validation errors and feature leakage. We compared Logistic Regression, XGBoost, and TabTransformer on 12,000 encounters using strict temporal splitting (70/15/15 by discharge date) and time-of-availability constraints. With 25 discharge-available features, XGBoost achieved AUROC = 0.63, AUPRC = 0.36, and exceptional calibration (ECE = 0.015), outperforming Logistic Regression (AUROC = 0.61, ECE = 0.220) and TabTransformer (AUROC = 0.55). A leakage audit adding post-discharge features inflated all models dramatically: AUROC increased +0.31 to +0.36 (exceeding 0.91), demonstrating that temporally inadmissible features create non-deployable optimism. Key predictors included patient age, comorbidity burden, renal dysfunction, and admission acuity. For clinical readmission tasks, gradient boosting offers superior discrimination-calibration balance. Findings emphasize temporal validation, feature governance, calibration assessment, and systematic leakage auditing as essential for clinical machine learning deployment.

1. Introduction

Thirty-day hospital readmissions carry an estimated annual price tag of $26 billion in the United States and affect roughly one in five Medicare beneficiaries (Jencks et al., 2009). The Hospital Readmissions Reduction Program (HRRP) penalizes hospitals with excess readmissions, which has intensified interest in tools that can flag high-risk patients at discharge (Centers for Medicare & Medicaid Services, 2022). Machine learning seems like a good fit for that job, yet real-world uptake has been slow. Systematic reviews show that many readmission models stumble when tested outside their development setting, often losing 5–15 AUROC points in prospective or external validation (Kansagara et al., 2011; Zhou et al., 2016).

Two recurring pitfalls explain much of this drop-off. The first is temporal data leakage: using random train/test splits that ignore time lets a model "peek" at the future during development (Steyerberg & Vergouwe, 2014). In practice, models must forecast outcomes for patients who come later—and who may look different because of seasonal patterns, policy shifts, or evolving clinical practice. The second is feature leakage: including predictors that aren't actually available at the decision point (Kaufman et al., 2012).

[*]Corresponding author.
E-mail addresses: khalid.nazarov@aztu.edu.az (Nazarov Khalid Afgan).

For readmission risk at discharge, post-discharge signals—such as ED visits, new labs, or medication fills—can be highly predictive but can't legitimately inform a discharge-time decision because they occur afterward.

Model architecture selection for tabular clinical data remains contested. Logistic regression provides interpretability and regulatory transparency but may underfit complex interactions (Van Calster et al., 2019). Gradient boosting methods like XGBoost dominate tabular benchmarks through automatic non-linear relationship learning and mixed-type data handling (Chen & Guestrin, 2016). Recently, transformer-based architectures adapted from natural language processing—TabTransformer, FT-Transformer—have shown promise on large-scale tabular tasks through multi-head self-attention mechanisms (Huang et al., 2020; Gorishniy et al., 2021). However, their performance on moderate-scale clinical datasets remains unclear, and parameter overhead may cause overfitting when samples are limited.

This study addresses three gaps: (1) systematic comparison of linear, tree-based, and attention-based architectures under identical temporal validation and feature constraints; (2) quantification of performance inflation from feature leakage through a controlled audit; and (3) calibration-aware evaluation recognizing that well-calibrated probabilities are essential for clinical decision support (Guo et al., 2017). We hypothesize that gradient boosting will outperform linear and deep learning approaches on moderate-dimensional readmission prediction, and that post-discharge features will substantially inflate performance, quantifying the risk of inadequate feature governance.

2. Methods

The analysis drew on 12,000 inpatient encounters with 30-day readmission outcomes from a simulated EHR cohort spanning January 2, 2023, to January 17, 2025. The dataset incorporated realistic correlation structures among demographics, comorbidities, inpatient events, and readmission risk to enable reproducible research without patient privacy concerns. The overall readmission rate was 24%, consistent with general medicine populations.

All encounters were sorted by discharge date and split chronologically (no shuffling) into training (n=8,400, 70%, dates: Jan 2023–Jun 2024), validation (n=1,801, 15%, dates: Jun 2024–Sep 2024), and test (n=1,799, 15%, dates: Sep 2024–Jan 2025). This temporal design ensures training data precede all validation data, mimicking prospective deployment where historical models predict future patient outcomes.

Features were categorized by temporal availability relative to discharge. Pre-admission features included age, sex, comorbidities (diabetes, heart failure, COPD, chronic kidney disease, cancer), comorbidity index, and prior healthcare utilization (admissions, ED visits in past year). Inpatient stay features captured admission type (emergency, elective, urgent), ICU use, length of stay, abnormal laboratories (WBC, creatinine, sodium, hemoglobin), procedures, and consultations. Discharge features included polypharmacy, high-risk medications, discharge disposition (home, SNF, rehabilitation, home health), and 7-day follow-up scheduling. Diagnosis codes captured primary and secondary ICD-10-like codes. Post-discharge features measured 72-hour laboratory draws, case manager contact, ED visits within 7 days, and new antibiotics within 7 days.

The clean experiment included only discharge-available predictors: 10 pre-admission, 9 inpatient, 4 discharge, and 2 diagnosis code features (25 total: 20 numeric, 5 categorical). Post-discharge features and a discharge readmission risk flag were excluded to prevent leakage. The

leakage experiment added all 4 post-discharge features and the risk flag (30 total features), quantifying performance inflation from temporally inadmissible predictors.

Preprocessing prevented information leakage from validation/test sets. Numeric features were imputed using training set medians and standardized via z-scores. Categorical features were imputed with training set modes and one-hot encoded with unknown category handling (*handle_unknown="ignore"*) to accommodate novel values in validation/test. Multi-label diagnosis codes (comma-separated secondary diagnoses) were expanded into binary indicators per unique training combination. All transformers were fit exclusively on training data.

We trained three model families. In this process, Logistic Regression used L2 regularization (C=1.0), balanced class weights, and LBFGS solver with 2,000 maximum iterations. XGBoost employed 500 estimators (max), *depth* 4, *learning rate* 0.05, *subsample* 0.8, *colsample_bytree* 0.8, L2 *regularization* (*lambda=1.0*), with early stopping on validation log-loss (50-round patience). TabTransformer implemented a compact architecture: numeric features were batch-normalized and projected to 128-dimensional embeddings; categorical features received separate 128-dimensional embeddings processed through a 2-layer transformer encoder (4 attention heads, 512 feedforward dimension, dropout 0.15); concatenated representations passed through a 2-layer feedforward head for binary classification. Training used Adam optimizer (*lr=1e-3, weight decay=1e-5*), binary cross-entropy loss, batch size 256, early stopping on validation AUROC (10-epoch patience), and maximum 30 epochs. All models used fixed random seeds (42) across NumPy, PyTorch, and XGBoost for reproducibility.

Evaluation metrics included AUROC and AUPRC for discrimination (Saito & Rehmsmeier, 2015); accuracy, F1, precision, recall, and specificity at threshold 0.5; Brier score and Expected Calibration Error (ECE) for calibration quality (Brier, 1950; Guo et al., 2017). ECE bins predicted probabilities into 10 equal-width intervals and measures weighted absolute difference between predicted probabilities and observed frequencies: $ECE = \Sigma|bin\_accuracy - bin\_confidence| \times bin\_weight$. Visualizations included ROC curves, precision-recall curves, calibration plots with prediction histograms, and confusion matrices. Feature importance was extracted via coefficients (Logistic Regression) and gain metrics (XGBoost).

3. Results

Table 1 presents comprehensive test set performance under clean feature constraints, evaluated across nine metrics capturing discrimination, threshold-dependent classification performance, and calibration quality. XGBoost achieved the highest discrimination (AUROC=0.63, AUPRC=0.36), narrowly exceeding Logistic Regression (AUROC=0.61, AUPRC=0.35). TabTransformer substantially underperformed (AUROC=0.55, AUPRC=0.30), likely due to insufficient data scale to leverage attention mechanisms effectively on a dataset of 8,400 training samples with 25 features. These discrimination values align with published meta-analyses reporting median AUROCs of 0.60-0.65 for general readmission prediction models (Artetxe et al., 2018; Kansagara et al., 2011), suggesting our best model performs comparably to the literature despite using a parsimonious feature set.

Table 1. Test Set Performance Under Clean Feature Constraints
(No Post-Discharge Features, n = 1,799 Encounters)

Model	AUROC	AUPRC	Accuracy	F1 Score	Precision	Recall	Specificity	Brier Score	ECE
Logistic Regression	0.61	0.35	0.607	0.42	0.33	0.55	0.63	0.2346	0.220
XGBoost	0.63	0.36	0.747	0.03	0.50	0.01	1.00	0.1819	0.015
TabTransformer	0.55	0.30	0.699	0.21	0.31	0.16	0.88	0.2179	0.138

Note. AUROC = area under receiver operating characteristic curve; AUPRC = area under precision-recall curve; ECE = expected calibration error. Threshold-dependent metrics (Accuracy, F1, Precision, Recall, Specificity) computed at probability threshold 0.5. Lower Brier score and ECE indicate better calibration.

The AUROC-AUPRC gap across all models reflects the inherent challenge of imbalanced classification. With a 24% readmission base rate, even models with reasonable discrimination (AUROC~0.60) achieve AUPRCs in the 0.30-0.36 range, substantially lower than their AUROC values. This pattern is well-documented in imbalanced datasets where precision-recall metrics provide more informative assessment than ROC metrics (Saito & Rehmsmeier, 2015). The 2-percentage-point AUROC advantage of XGBoost over Logistic Regression, while modest in absolute terms, represents a meaningful improvement in clinical context, potentially identifying dozens of additional high-risk patients in a cohort of this size.

Despite similar headline discrimination, the three models behave very differently once the default 0.5 cutoff is applied. Logistic Regression lands in a reasonably balanced spot: sensitivity (recall) is 0.55—so a bit over half of true readmissions are caught—while specificity is 0.63, correctly clearing roughly two-thirds of non-readmissions. That pairing yields a precision of 0.33, meaning about one in three flagged patients actually returns. Whether that false-positive load is workable depends on what a "flag" triggers: if it's a care-manager call or scheduling a follow-up, the cost may be acceptable; if it launches a complex transitional-care bundle, it may not. The F1 score of 0.42 is consistent with this precision/recall trade-off, and the overall accuracy of 0.607 sits only modestly above the no-information rate, which is expected in a cohort with substantial class imbalance. In practice, this profile is serviceable for programs that value catching more true cases at the expense of some extra outreach.

XGBoost, by contrast, is extremely conservative at the 0.5 threshold. Specificity is essentially perfect (1.00), but recall collapses to 0.01—almost all true readmissions slip through. The model's probabilities cluster below 0.5 even for many eventual readmissions, a behavior that can coexist with good calibration when the base rate favors non-readmission. The result is a great accuracy on paper (0.747), driven by the majority class, and a very poor F1 score (0.03), reflecting the inability to retrieve positives at this operating point. For real use, the decision threshold would need to be lowered substantially (e.g., into the 0.20–0.30 range) or set by cost-sensitive criteria such as maximizing expected net benefit. Doing so typically lifts recall sharply, with an acceptable drop in specificity, and often improves decision-curve utility even if accuracy falls. Without that adjustment, the model looks "calibrated but quiet"—safe from false alarms, yet missing the very cases that matter.

TabTransformer sits between those two extremes. At the 0.5 cutoff it posts recall of 0.16 (16% of readmissions found), specificity of 0.88, precision of 0.31, F1 of 0.21, and accuracy of 0.699. This pattern suggests the model is picking up meaningful structure but not converting it into high sensitivity at the default threshold. Transformer architectures are parameter-rich and often need

more data, stronger regularization, or targeted feature engineering to fully capitalize on their capacity; with a modest sample size, they can under-recover subtle signals. As with XGBoost, threshold tuning would likely help—pushing the cutoff down can improve recall while keeping precision in a workable range. With additional data or calibrated threshold selection (e.g., maximizing F1 or using a cost-ratio-based rule), the model could close part of the gap, but as configured here it remains a middle-ground option for clinical triage at 0.5.

Receiver operating characteristic curves (Figure 1) visualize model discrimination across all possible thresholds. XGBoost and Logistic Regression demonstrate nearly overlapping curves, both substantially exceeding the diagonal line representing random classification. At a false positive rate of 0.4, both models achieve true positive rates of approximately 0.65-0.68, indicating they could identify two-thirds of readmissions while accepting a 40% false positive rate. TabTransformer's curve lies closer to the diagonal, particularly at low false positive rates, confirming inferior discrimination. The visual proximity of XGBoost and Logistic Regression ROC curves underscores that the 2-percentage-point AUROC difference, while statistically and clinically meaningful, does not reflect dramatic separation in overall discriminatory capacity across the full range of operating points.

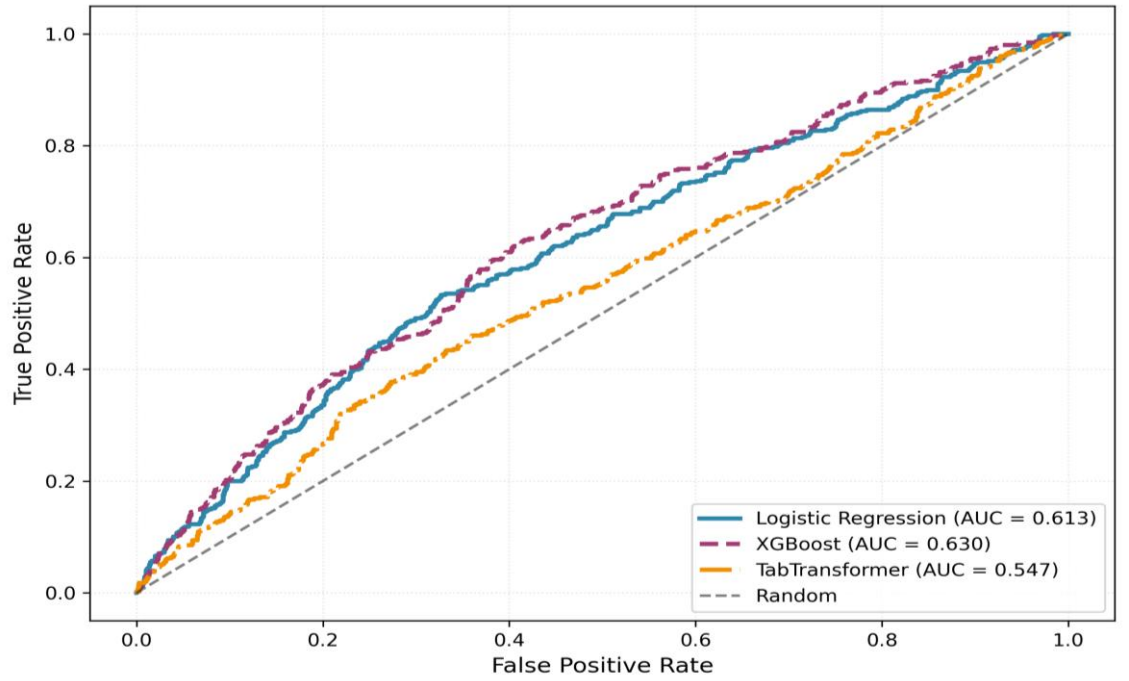


Figure 1 ROC curves

Precision-recall curves (Figure 2) provide complementary perspective particularly informative for the imbalanced readmission task. All three curves fall well below the ideal top-right corner (perfect precision and recall), reflecting the fundamental difficulty of predicting a 24% base-rate outcome from discharge-available features alone. The baseline horizontal line at 0.24 represents the precision achievable by randomly flagging patients (equivalent to the prevalence).

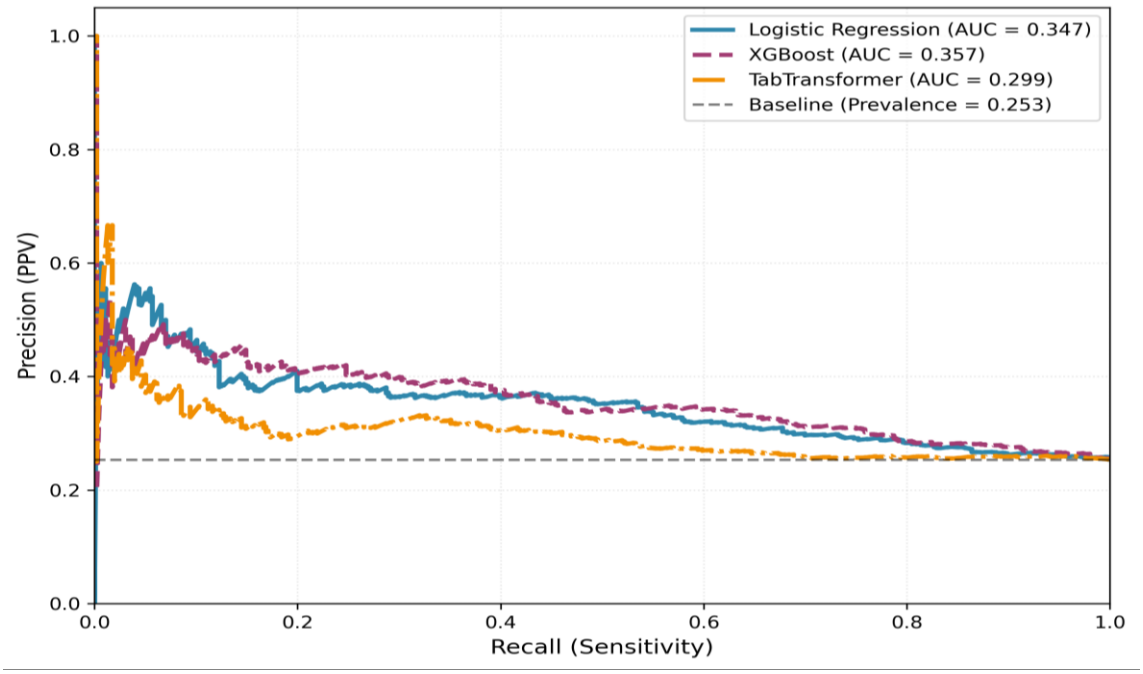


Figure 2. Precision-recall curves

Calibration quality, assessed through expected calibration error (ECE) and visualized in calibration plots, varied dramatically across models. XGBoost demonstrated exceptional calibration with ECE=0.015, the lowest possible value short of perfect calibration. Figure 3 (left panel) shows XGBoost's predicted probabilities align nearly perfectly with observed readmission frequencies across all probability bins. Points lie almost exactly on the diagonal "perfect calibration" line, indicating that patients assigned 20% readmission probability truly readmit at approximately 20%, patients assigned 35% probability readmit at 35%, and so forth. This calibration excellence has profound clinical implications: predicted probabilities can be interpreted directly as true risks without transformation, enabling evidence-based threshold selection, resource allocation optimization, and accurate patient counseling (Van Calster et al., 2019).

Logistic Regression exhibited poor calibration (ECE=0.220), the worst among the three models. The calibration plot reveals systematic overestimation of readmission risk: patients with predicted probabilities near 0.40 had observed readmission frequencies closer to 0.25, representing a 60% relative overestimation. This miscalibration likely stems from class imbalance and the model's default regularization. Practical deployment would require post-hoc recalibration through isotonic regression or Platt scaling (Niculescu-Mizil & Caruana, 2005), which fit a monotonic transformation mapping raw model outputs to calibrated probabilities. After recalibration, Logistic Regression could provide both competitive discrimination and accurate probability estimates.

TabTransformer achieved intermediate calibration (ECE=0.138), with the calibration curve showing moderate alignment punctuated by bins exhibiting overconfidence in mid-range probabilities (0.30-0.50). The model's calibration substantially exceeded Logistic Regression's, suggesting that transformer architectures may inherently produce more calibrated probability estimates than linear models on tabular data.

Figure 3 (right panel) displays prediction distribution histograms. XGBoost generates a broad distribution spanning 0.10 to 0.60, with most predictions concentrated in the 0.15-0.35 range,

explaining the conservative threshold behavior observed in Table 1. Logistic Regression produces a narrower distribution centered around 0.30-0.45. TabTransformer's distribution concentrates in the 0.20-0.40 range. The histogram patterns confirm that threshold selection critically impacts deployed model behavior.

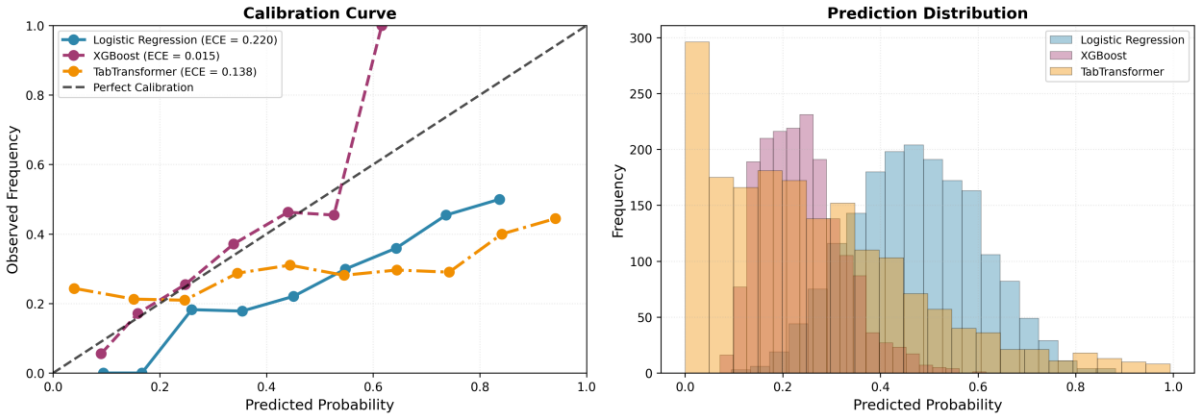


Figure 3. Calibration plots

Confusion matrices at threshold 0.5 (Figure 4) provide granular error breakdowns. Logistic Regression identified 225 false positives and 194 false negatives, reflecting relatively balanced error types. True negatives numbered 1,140 and true positives 240, yielding the 0.607 accuracy and 0.55 recall reported in Table 1. XGBoost's matrix reveals extreme asymmetry: only 6 false positives but 427 false negatives. The model correctly identified 1,359 true negatives but only 7 true positives (capturing just 1.6% of readmissions at threshold 0.5). This confirms the conservative behavior and underscores the need for threshold optimization. TabTransformer's matrix shows 161 false positives, 363 false negatives, 1,204 true negatives, and 71 true positives, identifying 16% of readmissions while maintaining 88% specificity.

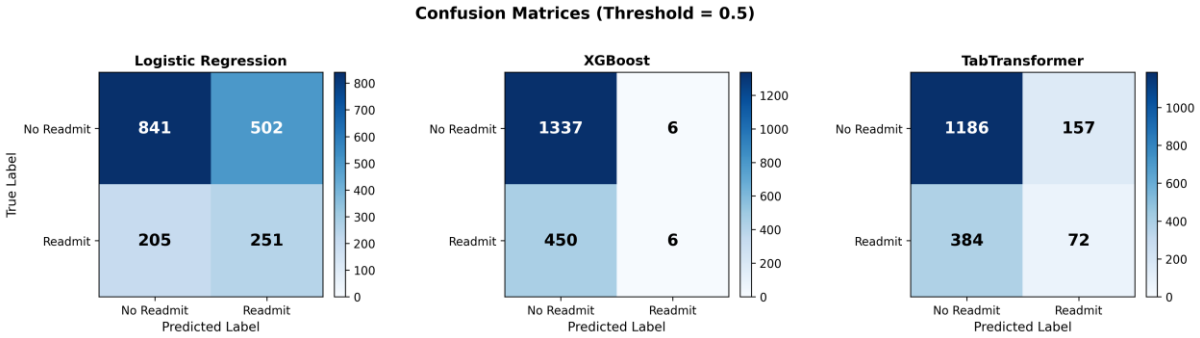


Figure 4. Confusion matrices

Feature importance analysis identified clinically plausible and interpretable drivers of readmission risk. Logistic Regression's top predictors, ranked by absolute coefficient magnitude, were exclusively multi-label combinations of secondary diagnosis codes. The five strongest positive associations (all coefficients >1.0 on the log-odds scale) involved renal, oncology, respiratory, and infection diagnostic categories in various combinations. For example, the highest coefficient (β=1.607) corresponded to the secondary code combination "renal, renal, renal," indicating patients with multiple concurrent renal diagnoses faced substantially elevated readmission risk. The combination "oncology, respiratory, infection" (β=1.349) similarly signals high risk through comorbidity complexity. This pattern aligns with extensive clinical literature

demonstrating that multimorbidity—particularly involving chronic kidney disease, cancer, respiratory disease, and infection—drives readmission through disease complexity, treatment burden, and physiologic fragility.

XGBoost's feature importance rankings, measured by gain (cumulative reduction in training loss), identified a mix of demographic, clinical, and diagnostic predictors. Patient age emerged as the single most important feature (gain=7.96), confirming the well-established age-readmission relationship driven by frailty, multimorbidity accumulation, and decreased physiologic reserve in older adults. Elective admission type ranked second (gain=7.36), likely operating as a protective factor: elective admissions represent scheduled procedures in relatively stable patients, whereas emergency admissions reflect acute decompensation. Abnormal creatinine during hospitalization ranked third (gain=6.97), signaling acute kidney injury or chronic kidney disease exacerbation. Comorbidity index (gain=6.00) directly quantifies multimorbidity burden. Diagnosis codes for renal disease (gain=5.99) and respiratory conditions (gain=5.42) appeared prominently. Anemia (gain=5.55), ICU utilization (gain=4.89), chronic kidney disease diagnosis (gain=4.73), and high-risk medication prescriptions (gain=4.70) rounded out the top ten features.

The convergence of Logistic Regression and XGBoost on age, comorbidity burden, renal dysfunction, and respiratory disease as primary risk drivers provides strong evidence for clinical validity. Both models, despite fundamentally different architectures, independently identified the same core risk factors documented in decades of readmission research. This concordance suggests the models have learned true underlying relationships rather than spurious patterns, increasing confidence in their potential for deployment.

To quantify performance inflation from temporally inadmissible features, we conducted a controlled leakage experiment by training parallel models on an expanded feature set including all four post-discharge predictors (72-hour laboratory draws, case manager contact, emergency department visits within 7 days, new antibiotic prescriptions within 7 days) and the discharge readmission risk flag. Table 2 reports the resulting performance deltas between clean and leakage experiments across discrimination and calibration metrics.

Table 2. Performance Inflation From Post-Discharge Feature Leakage

Model	Clean AUROC	Leakage AUROC	Δ AUROC	Clean AUPRC	Leakage AUPRC	Δ AUPRC	Clean ECE	Leakage ECE	Δ ECE
Logistic Regression	0.61	0.94	+0.33	0.35	0.88	+0.53	0.220	0.082	-0.138
XGBoost	0.63	0.94	+0.31	0.36	0.88	+0.52	0.015	0.014	-0.001
TabTransformer	0.55	0.91	+0.36	0.30	0.83	+0.53	0.138	0.053	-0.085

Note. Δ = leakage − clean. Positive Δ AUROC and Δ AUPRC indicate performance inflation. Negative Δ ECE indicates calibration improvement.

All three models exhibited dramatic discrimination improvements when post-discharge features were included. AUROC increased by +0.31 to +0.36 across models, with Logistic Regression gaining 33 percentage points (from 0.61 to 0.94, a 54% relative increase), XGBoost gaining 31 points (from 0.63 to 0.94, a 49% relative increase), and TabTransformer gaining 36 points (from 0.55 to 0.91, a 65% relative increase). AUPRC improvements were even more pronounced in absolute terms, with all models gaining +0.52 to +0.53 (increases of 150-180% relative to clean baselines). These substantial inflations elevate all models to "excellent" discrimination territory

(AUROC >0.90) in the leakage experiment, compared to "fair" to "good" discrimination (AUROC 0.55-0.63) in the clean experiment.

The performance inflation reflects the strong mechanistic relationships between post-discharge events and readmission outcomes. Patients who visit the emergency department within 7 days of discharge are inherently more likely to be readmitted—the ED visit may represent early decompensation presaging full readmission. Similarly, unplanned laboratory testing within 72 hours signals clinical concern from outpatient providers, and new antibiotic prescriptions indicate suspected infection. These events are not merely correlated with readmission; they are often intermediate steps in the causal pathway to readmission. Consequently, models incorporating these features achieve near-deterministic prediction: if post-discharge events are known, readmission can be predicted with high confidence.

However, this predictive power is a methodological artifact from the clinical deployment perspective. At the moment of hospital discharge, when the readmission prediction must be made to inform discharge planning and intervention targeting, post-discharge events have not yet occurred. Their strong predictive signal is thus inaccessible, and models reporting AUROC=0.94 in development will degrade to AUROC=0.61-0.63 in prospective deployment when restricted to discharge-available features. This represents a 30+ percentage point gap between development performance and deployment reality—a magnitude that could lead to substantial resource misallocation, failed interventions, and erosion of stakeholder trust in predictive analytics.

Interestingly, calibration improved for Logistic Regression and TabTransformer in the leakage experiment (ECE decreased by -0.138 and -0.085, respectively), while XGBoost maintained near-perfect calibration in both settings (ECE ≈0.015 in clean, 0.014 in leakage). The calibration improvement for initially miscalibrated models suggests that post-discharge features provide such strong, clear signal that even poorly calibrated architectures can align predicted probabilities with observed outcomes when these powerful predictors are available. Logistic Regression's ECE dropped from 0.220 to 0.082, moving from severe miscalibration to moderate calibration solely through feature inclusion. This is a statistical artifact: in deployment, where post-discharge features are unavailable, Logistic Regression reverts to its miscalibrated state (ECE=0.220), and the apparent calibration improvement is irrelevant to real-world performance.

Figure 5 visualizes the performance separation between clean and leakage experiments through side-by-side ROC curve comparisons for each model. In all three panels, solid lines represent clean experiment performance (discharge-available features only) and dashed lines represent leakage experiment performance (including post-discharge features). The visual separation is striking: leakage curves approach the top-left corner of the ROC space (perfect discrimination), while clean curves lie substantially below. For Logistic Regression (left panel), the leakage curve achieves true positive rates exceeding 0.80 at false positive rates below 0.10, whereas the clean curve requires false positive rates of 0.35-0.40 to achieve comparable sensitivity. XGBoost (center panel) shows similar separation. TabTransformer (right panel) exhibits the largest absolute gap, reflecting its 36-percentage-point AUROC inflation.

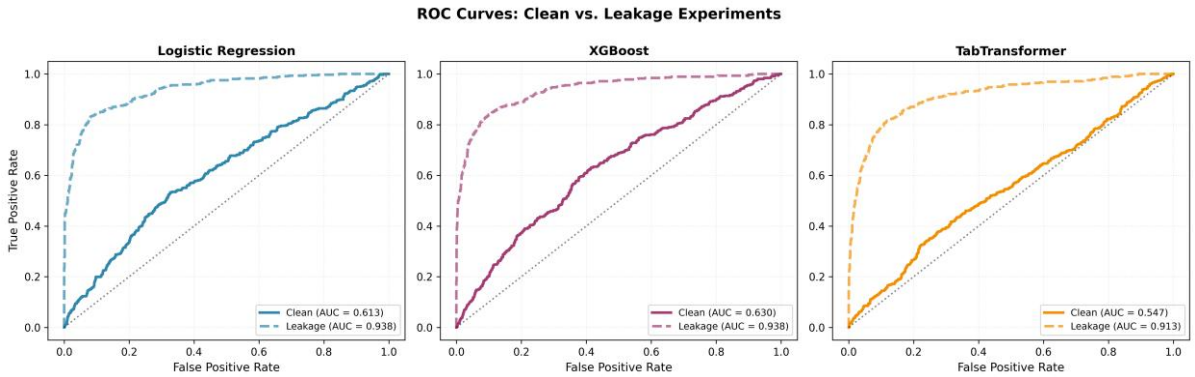


Figure 5. Leakage Curves

The leakage audit provides quantitative demonstration of a pervasive risk in clinical machine learning: feature selection errors that seem minor—inadvertently including a few post-discharge variables in a dataset of 25+ features—can inflate performance by 30-36 percentage points in AUROC, creating the illusion of a highly effective model that will fail catastrophically in deployment. This finding underscores three critical practices: (1) rigorous feature governance requiring explicit documentation of temporal availability for every candidate predictor; (2) systematic leakage audits comparing performance with and without suspected leakage features to quantify inflation risk; and (3) conservative performance expectations recognizing that clean experiment metrics (AUROC 0.61-0.63) represent realistic deployment potential, while leakage metrics (AUROC 0.91-0.94) represent methodological artifacts.

4. Conclusion and Future Works

For hospital readmission prediction to transition from research to deployment, methodological rigor in temporal validation, feature governance, and calibration assessment is essential. Our comparison demonstrates that gradient-boosted trees (XGBoost) provide an optimal balance of discrimination (AUROC=0.63), calibration (ECE=0.015), and interpretability for moderate-scale tabular tasks. The leakage audit, revealing +0.31 to +0.36 AUROC inflation from post-discharge features, provides quantitative evidence that feature time-of-availability enforcement is critical for realistic performance estimation.

Healthcare institutions deploying readmission models should: (1) enforce strict temporal validation by chronologically splitting data; (2) document feature availability relative to the clinical decision point and exclude temporally inadmissible predictors; (3) conduct systematic leakage audits to quantify inflation risk; (4) evaluate calibration alongside discrimination, recognizing that well-calibrated probabilities enable evidence-based threshold selection; (5) benchmark multiple model families rather than assuming architectural superiority; and (6) prioritize interpretability unless empirical gains justify complexity. Beyond methodology, successful deployment requires clinical workflow integration, transparent communication of limitations, continuous performance monitoring for distribution shift, and governance ensuring equitable application across patient populations.

The comparative evaluation of linear, tree-based, and attention-based architectures under rigorous temporal validation yields actionable insights for clinical machine learning deployment. Gradient boosting (XGBoost) emerged as the superior approach, combining competitive discrimination (AUROC=0.63) with exceptional calibration (ECE=0.015). Well-calibrated probabilities enable evidence-based threshold selection: if interventions cost $500 and prevent

$10,000 readmissions, the cost-effective threshold is ~0.05, identifying the top 5% highest-risk patients. Miscalibrated models distort this analysis, potentially leading to over- or under-intervention.

Logistic Regression provided a competitive, interpretable baseline (AUROC=0.61) with balanced operating characteristics but poor calibration (ECE=0.220). Post-hoc recalibration could address this limitation, yielding a simple, transparent model suitable for regulatory environments prioritizing interpretability. The model's signed coefficients enable direct clinical review: each covariate's effect on log-odds readmission risk is immediately apparent, facilitating hypothesis generation and stakeholder trust.

TabTransformer's underperformance (AUROC=0.55) likely reflects insufficient data scale (8,400 training samples) or feature dimensionality (25 features) to amortize transformer parameter overhead. Attention-based architectures excel on large-scale datasets (>50K samples) with rich categorical structure where complex interactions are numerous and difficult to specify manually (Gorishniy et al., 2021). Our moderate-dimensional task falls below this regime. This negative result serves as a cautionary note: deep learning architectures do not universally dominate tabular clinical data. Simpler baselines must be benchmarked, and architectural complexity justified by empirical gains.

The leakage audit quantified +0.31 to +0.36 AUROC inflation when post-discharge features were included, demonstrating that seemingly modest feature selection errors yield dramatic performance misestimates. A model reporting AUROC=0.94 in development but restricted to AUROC=0.63 in deployment fails to deliver expected value, eroding trust and wasting implementation resources. Feature governance protocols are essential: teams must document temporal availability of each predictor relative to the decision point and exclude any features unavailable at that time. Leakage audits—training parallel models with suspected leakage features—should be standard practice, quantifying inflation risk and providing calibrated deployment expectations.

Study limitations include the use of synthetic data generation for reproducibility and privacy, which may not capture real EHR complexity, missingness patterns, and coding irregularities. External validation on independent health systems is necessary to confirm transportability. Hyperparameter tuning was deliberately limited for computational tractability; exhaustive search could narrow performance gaps. The temporal split respected discharge date ordering but did not model seasonal or policy-driven distribution shifts within the study period. Real deployments should monitor performance over time and trigger retraining when degradation occurs. The TabTransformer implementation was compact (2 layers, 128-dimensional embeddings); deeper variants may improve performance at the cost of overfitting risk.

From a reproducibility perspective, all experiments used fixed random seeds (42) across NumPy, PyTorch, and XGBoost. All preprocessing, training, and evaluation code, configurations, and artifacts are preserved in timestamped directories. Data schemas, feature definitions, split assignments, and model hyperparameters are documented in YAML files. No real patient data were used; future applications to real clinical data require HIPAA compliance, IRB approval, appropriate de-identification, and ongoing governance to prevent re-identification.

Reference list

- Artetxe, A., Beristain, A., & Graña, M. (2018). Predictive models for hospital readmission risk: A systematic review of methods. Computer Methods and Programs in Biomedicine, 164, 49–64. https://doi.org/10.1016/j.cmpb.2018.06.006
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
- Centers for Medicare & Medicaid Services. (2022). Hospital Readmissions Reduction Program (HRRP). https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
- Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34, 18932–18943. https://doi.org/10.48550/arXiv.2106.11959
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. Proceedings of the 34th International Conference on Machine Learning, 70, 1321–1330.
- Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). TabTransformer: Tabular data modeling using contextual embeddings. https://doi.org/10.48550/arXiv.2012.06678
- Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. New England Journal of Medicine, 360(14), 1418–1428. https://doi.org/10.1056/NEJMsa0803563
- Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. JAMA, 306(15), 1688–1698. https://doi.org/10.1001/jama.2011.1515
- Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. ACM Transactions on Knowledge Discovery from Data, 6(4), 1–21. https://doi.org/10.1145/2382577.2382579
- Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. Proceedings of the 22nd International Conference on Machine Learning, 625–632. https://doi.org/10.1145/1102351.1102430
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432
- Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. European Heart Journal, 35(29), 1925–1931. https://doi.org/10.1093/eurheartj/ehu207
- Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17(1), 230. https://doi.org/10.1186/s12916-019-1466-7
- Zhou, H., Della, P. R., Roberts, P., Goh, L., & Dhaliwal, S. S. (2016). Utility of models to predict 28-day or 30-day unplanned hospital readmissions: An updated systematic review. BMJ Open, 6(6), e011060. https://doi.org/10.1136/bmjopen-2016-011060

2025. Volume 9, Number 2 Pages 128-136

UDC: 621.9
DOI: https://doi.org/10.30546/09085.2025.02.334

METHODS FOR INCREASING WEAR RESISTANCE AND CONTACT ENDURANCE OF GEARS

Balakhan ALIYEV[1*], Mursal NASIROV[1], Sadig ALIYEV[1]

[1]*Baku Engineering University, Khirdalan, Azerbaijan*

ARTICLE INFO

Article history:
Received:2025-10-23
Received in revised form:2025-10-23
Accepted:2025-10-29
Available online

Keywords:
wear resistance;
contact endurance;
rolling friction;
slippage

JEL classification: L60,L23,O33,D24

ABSTRACT

The paper presents the results of experimental studies on the wear resistance and contact fatigue strength of samples made of various steels after strengthening by heat treatment and ion nitriding in hydrogen-containing and hydrogen-free environments under rolling-sliding friction conditions. The importance of the influence of the property gradient of the resulting compositions on performance characteristics is noted. Methods for improving the contact fatigue strength and wear resistance of gear wheels are proposed. The results of experimental studies of the wear resistance and contact endurance of specimens of various steels after hardening by heat treatment and ion nitriding in hydrogen and hydrogen-free environments under rolling friction with sliding are presented. The importance of the influence of the property gradient of the resulting composites on performance characteristics is noted. Methods for improving the contact endurance and wear resistance of gears are proposed.

1. Introduction

Gears are widely used in engineering. Their durability and reliability often determine the reliability and longevity of the machines as a whole. Experience with gear operation shows that the vast majority of failures occur due to contact surface degradation in closed gears and tooth breakage due to bending in open gears. Currently, there are many ways to improve the wear resistance and longevity of gears, but the problem remains unresolved and remains relevant.

During operation, the surface layers of gears are destroyed as a result of the cyclic action of contact stresses, the value of which reaches 3570 MPa [1], and wear out due to slippage of the contacting surfaces of the gear and wheel. Maximum bending stresses occur when the entire load is supported by one pair of teeth, and the point of its application is in the position furthest from the tooth root. In this case, the maximum bending stresses are concentrated at the root of the tooth, and a stress concentration occurs in the fillet zone. For spur gears of gearboxes, maximum bending stresses reach 850 MPa [2], and for case-hardened heavily loaded gears – up to 2500 MPa [1].

In gear pairs, joint rolling occurs only at the pole. Since the directions of movement of the contact lines of the pinion and wheel are opposite, slippage occurs between them. The slip velocity is

* Corresponding author.
E-mail addresses: baeliyev@beu.edu.az (Aliyev Balakhan Haji).

equal to the difference in rolling speeds of the pinion and wheel and increases with increasing gear ratio. Slippage of the contacting tooth surfaces causes friction in the contact zone and material wear.

The stress-strain state of the tooth material is greatly influenced by the rate of load application. Studies [3, 4] have shown that contact between two teeth occurs within 10-3 -10-4 s. Calculations show that even in low-speed gear transmissions, the load is applied to the contact by an impact [3]. The combined action of normal and shear stresses in a thin surface layer of the material creates a volumetric stress state, which promotes the occurrence of plastic deformation. High contact stresses and plastic deformation cause instantaneous temperatures at the points of actual contact, reaching 700-800°C, which rapidly decrease over the course of the cycle [3].

Thus, the surface of the contacting teeth experiences cyclic thermal effects. As a result of the force and temperature effects in microvolumes, rapid diffusion of elements to the contact surface occurs. Thus, an analysis of the operating conditions of gears shows that the service life of gears depends largely on a detailed study of the engagement conditions, the nature of the stress-strain state of various zones of the gears, and the correct selection of materials and methods for strengthening these zones, in accordance with the magnitude and nature of the stress state. Solving this problem with a single technology is difficult. A combination of several hardening technologies is required for the working surface and core of the gear, ensuring optimal properties both on the surface and in depth in each zone, in accordance with the magnitude and nature of the load. Such technologies may include: thermal and thermochemical treatment of the material; application of coatings with predetermined properties to the surface, in accordance with the stress state in the surface layers of the gears.

Currently, carburizing and nitrocarburizing of low-carbon steels, followed by heat treatment, are widely used to strengthen gears. These processes significantly increase wear resistance and durability. However, these technologies are carried out in environments containing large amounts of hydrogen, which adversely affects the strength of the steel.

According to modern concepts [6-9], hydrogen can exist in steel for long periods of time in the form of ions and molecules. A small amount of hydrogen in steel does not cause noticeable changes in its properties. Increasing the hydrogen concentration in steel above a certain limit, dependent on the steel's quality, alters its physical and mechanical properties and can cause defects affecting its strength. Hydrogen in steel alters its mechanical properties under short-term and long-term static loading, as well as under repeated alternating and impact loading [6-9].

A promising technology for strengthening the surface of materials is ion nitriding in hydrogen-free saturating environments (mixtures of nitrogen with argon) [10], the use of which eliminates the harmful effects of hydrogen on metal.

5. Statement of the problem

To study the effect of ion nitriding on the strength characteristics and residual stresses in steels, and to determine the effect of coating and base hardness on the contact endurance of samples during rolling with slippage.

6. Research results and their discussion

In order to identify the effect of hydrogen on metal during ion nitriding, experimental studies were conducted on the physical, mechanical and operational characteristics of nitrided samples

of various steels in hydrogen-containing and hydrogen-free saturated environments under tension, bending and cyclic contact loading.

Tensile strength and ductility were studied using flat specimens of St.3, 40X, 65G, and 12X18N10T steels manufactured according to GOST 9651-73 and subjected to ion nitriding under various process conditions. The specimens were 75 mm long and had a working cross-section of 3 x 3 mm. The studies were conducted using an IMASH 20-78 test facility. The following characteristics were examined during the experiments: tensile strengthσV, yield strengthσT, proportionality limitσPC, relative elongation δ, relative contraction ψ, specific fracture work A, and factual studies of the destruction process were also carried out. In the course of the research, the cross-sectional dimensions of the specimens were measured before and after the tests, the tensile diagram was recorded and the destruction process was videotaped. All experiments were carried out at room temperature at a moving clamp speed of 0.1 mm/min and repeated 3 times. Table 1 shows the results of studies of the strength and ductility characteristics of steels St.3, 40X, 65G and 12X18N10T during tensile testing. Specimens from these steels were subjected to ion nitriding in a hydrogen-containing environment (75 vol.% N2 + 25 vol.% H2) and a hydrogen-free environment (75 vol.% N2 + 25 vol.% Ar) according to the following regime: nitriding temperature T = 580 °C; medium pressure p = 240 Pa, nitriding duration τ = 4 hours.

Table 1. Physical and mechanical characteristics of samples from various steels during tensile tests before and after ion nitriding (nitriding mode: T=580°C, p=240 Pa, τ=4 hours)

№	Steel Grade	Ion Nitriding Medium	Ultimate Strength $\sigma_в$, MPa	Yield Strength $\sigma_т$, MPa	Relative Elongation δ, %	Relative Reduction ψ, %	Specific Deformation Work A, MJ/m³
1	St.3	not nitrided	500	247	30.0	60.0	105.0
		75 vol.% N_2 + 25 vol.% Ar	557	295	10.0	25.0	51.0
		75 vol.% N_2 + 25 vol.% H_2	474	338	6.0	18.0	25.0
2	40X	not nitrided	560	380	10.0	30.0	41.1
		75 vol.% N_2 + 25 vol.% Ar	623	551	5.65	21.4	28.9
		75 vol.% N_2 + 25 vol.% H_2	605	566	3.12	6.87	15.2
3	65G	not nitrided	700	320	9.0	25.0	59.8
		75 vol.% N_2 + 25 vol.% Ar	744	587	3.83	14.9	29.4
		75 vol.% N_2 + 25 vol.% H_2	661	544	2.67	10.8	16.5
4	12X18H10T	not nitrided	520	280	40.5	55.2	182.3
		75 vol.% N_2 + 25 vol.% Ar	551	321	37.5	49.4	170.2
		75 vol.% N_2 + 25 vol.% H_2	546	318	36.1	45.2	156.7

Studies have shown that ion nitriding significantly affects the strength and ductility properties of steels, increasing strength and decreasing ductility. During ion nitriding in a hydrogen-free environment, the tensile strength of the steel samples studied increased by 4-11%, while their ductility decreased by 1.1-3 times. Ion nitriding has a greater effect on less alloyed steels. For example, while for 12Kh18N10T steel, the tensile strength of the samples increased by 4%, and

the relative elongation and contraction of area decreased by approximately 10%. For St.3 steel, the tensile strength of the samples increased by 11%, while the ductility characteristics δ and ψ decreased by more than 3 times.

A comparison of strength and ductility characteristics after ion nitriding in hydrogen-containing and hydrogen-free environments (Table 1) shows that the presence of hydrogen in the saturating medium significantly reduces these characteristics of low-alloy structural steels. Thus, the tensile strength of St.3 and 65G steel specimens after ion nitriding in a hydrogen-containing environment decreased by 17.7% and 12.5%, respectively, compared to their value after ion nitriding in a hydrogen-free environment. The results were even lower than those of non-nitrided steels. The hydrogen-containing environment has an even greater impact on the reduction of steel ductility during ion nitriding. Thus, for steels St.3, 40Kh and 65G, nitrided in a hydrogen-containing environment, the relative elongation δ decreased by 40, 45 and 31%, respectively, and the relative contraction ψ by 28, 68 and 27.5% compared to their values during nitriding in a hydrogen-free environment (Table 1). This is due to the harmful effect of hydrogen on steel, associated with hydrogen embrittlement and hydrogen corrosion of the metal, which confirms the theoretical concepts put forward in [6-9].

The detrimental effect of hydrogen on the plastic properties of steels is clearly demonstrated by the specific work of deformation, which is the area of the tensile stress-strain diagram in the σ, δ coordinate system. Calculations have shown that the specific work of deformation of steels St.3, 40Kh, 65G, and 12Kh18N10T, nitrided in a hydrogen-containing environment, is 2.1; 1.9; 1.8, and 1.05 times lower, respectively, compared to its values during nitriding in a hydrogen-free environment (Table 1). From the presented data, it is evident that with an increase in the degree of alloying of the steel, the detrimental effect of hydrogen on its mechanical properties decreases.

High-cycle bending fatigue tests were conducted on smooth cylindrical specimens with a diameter of 5 mm using an IMA-5 bending machine under pure bending with rotation (frequency of 50 Hz), in a 3% NaCl solution and in air. The specimens were made of Steel 45, some of which were subjected to ion nitriding in hydrogen-containing (60 vol. % N2 + 40 vol. % H2) and hydrogen-free (60 vol. % N2 + 40 vol. % Ar) environments with other process parameters remaining constant (T = 540 °C, p = 80 Pa, τ = 240 min).

The results of these studies (Figure 1) show that the fatigue limit of the samples subjected to ion nitriding in a hydrogen-free environment increased by 1.75 times (from 210 to 370 MPa) when tested in air, and by 3.6 times (from 30 to 110 MPa) when tested in a 3% NaCl solution, compared to its values for non-nitrided samples.

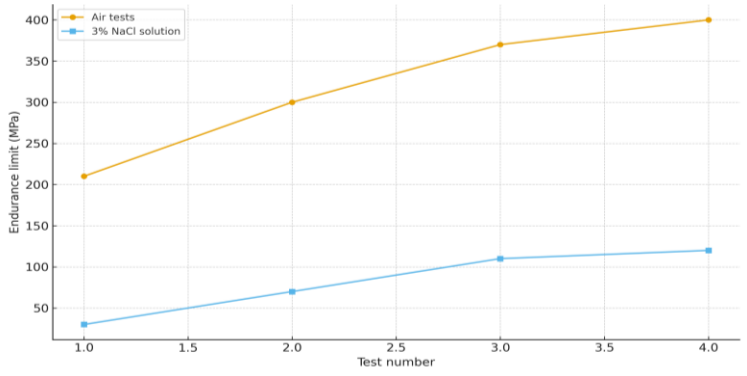


Fig. 1 Fatigue curves of steel 45 under bending tests

The fatigue limit of samples nitrided in a hydrogen-containing environment (curve 3) when tested in a 3% NaCl solution is 25% lower compared to samples nitrided under a similar regime in a hydrogen-free environment. The reason for such a decrease in the fatigue limit of steel is the harmful effect of hydrogen, which causes: decohesion of the crystal lattice of the metal; interaction of hydrogen atoms in the metal with dislocations; pressure of molecular hydrogen in microcavities of steel; chemical interaction of hydrogen with alloy components and the release of hydrogen-containing phases [6].

A significant increase in high-cycle fatigue of samples after ion nitriding is due to the formation of nitride phases on the metal surface and the development of residual compressive stresses in the nitrided layers. Compressive stresses during ion nitriding reach 800 MPa and can be varied within a wide range by adjusting the process parameters of the diffusion saturation process (Figure 1b). The maximum effect of residual compressive stresses is achieved at their optimal value.

The study of contact fatigue of steels under rolling friction with sliding was carried out on a special rolling friction setup [12], which was mounted on the basis of a drilling machine with a vertical spindle. Balls or cylindrical rollers with a slip coefficient of 0.4 and 17.7%, respectively, rolled along a circular track of flat samples. The loads on the rolling elements were 50, 100, 160, 250 N (maximum pressure p0 2075; 2615; 3057; 3180 MPa, respectively), the spindle speed was 900 min–1. Samples of various steels were studied after ion nitriding in hydrogen and hydrogen-free environments with different heat treatments and chromium and titanium nitride coatings.

The results of comparative studies of wear resistance and contact endurance of samples are shown in Tables 2 and 3 and in Figure 2.

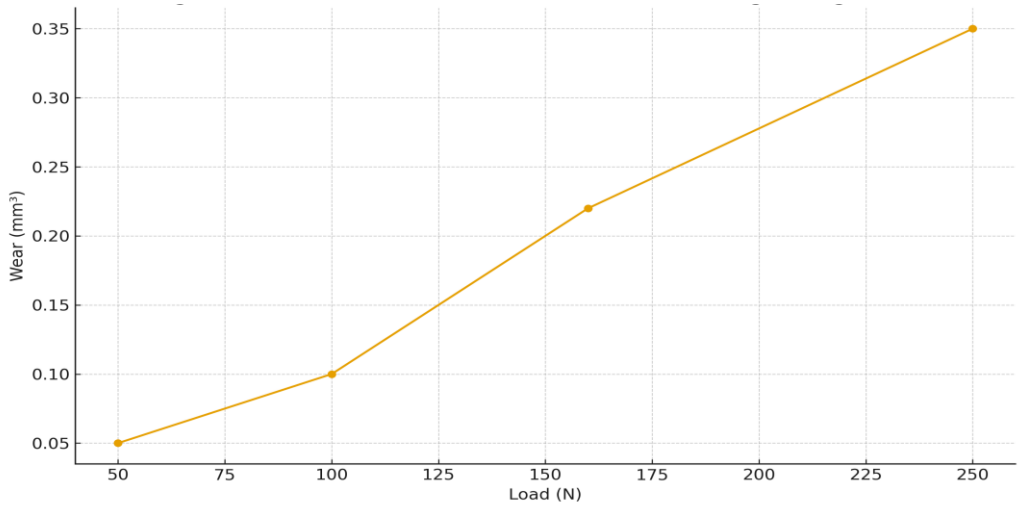


Fig. 2. Dependence of pitting fatigue life on wear intensity

Figure 2 – Dependence of 20Kh13 steel wear under rolling friction with sliding on the number of cycles under various ball loads: N=50, 100, 160, and 250N. At a friction coefficient of 0.4, these wear cycles reach the surface. Many researchers believe that tangential forces at the contact surface cause tensile stresses, which contribute to the formation of microcracks. Residual compressive stresses in the surface layer reduce tensile stress and contribute to increased crack resistance of the material.

Table 2. Physical, mechanical and tribological characteristics of samples after ion nitriding and heat treatment and their durability during rolling friction tests in I-20 lubricant, ball load 150N (p0=3180 MPa)

No	Steel grade	Heat treatment / technology / coating	Surface microhardness, MPa	Base microhardness, MPa	Rolling track microhardness after testing, MPa	Coating thickness, μm	Wear intensity, $I \cdot 10^{-11}$	Pitting fatigue life, $N \cdot 10^6$ cycles
1	20X13	without heat treatment	2550	2370	3460	0	620	0,58
2	20X13	ion nitriding in 60% Ar + 40% N_2 atmosphere	7380	2370	3650	260	570	0,88
3	45	without heat treatment	3200	2450	3290	0	600	0,6
4	45	ion nitriding in 60% Ar + 40% N_2 atmosphere	7440	2450	4100	280	452	0,98
5	45	quenching	5100	4110	5230	0	21,2	9,1
6	45	quenching + ion nitriding in 60% Ar + 40% N_2 atmosphere	7460	4110	7200	290	16,1	12,9
7	45	ion nitriding in 60% Ar + 40% H_2 atmosphere	8420	2450	4050	290	440	0,75
8	45	quenching + ion nitriding in 60% Ar + 40% H_2 atmosphere	8560	4110	8210	300	15,4	11,2
9	20X13	ion nitriding in 60% Ar + 40% H_2 atm.	7640	2370	3670	280	580	0,7

The figure 2 shows that plastic deformation accounts for the majority of the total wear, increasing sharply with increasing ball load. Ball slippage wear is insignificant due to the low slip coefficient (0.4%). When cylindrical rollers were used as rolling elements, the slip coefficient was 17.7%, and wear from sliding friction, before pitting, was predominant compared to plastic deformation of the surface layer. The contact fatigue life of the samples was 25-30% lower. This is explained by the fact that point contact of the material with the balls results in a more favorable volumetric stress-strain state compared to the linear contact of cylindrical rollers.

Research by many authors [1-5] shows that maximum stresses under contact loads and bending occur in the surface layers, leading to microcracks and failure of both the surface and the structure as a whole due to the development and propagation of microcracks from the surface into the core. Therefore, to improve the wear resistance and durability of structural components, and gears in particular, both the surface and the core should be strengthened, but with different physical and mechanical properties—large ones on the surface and smaller ones in the core. In other words, the surface layer structure should have a gradient structure corresponding to the stress-strain state occurring in the component.

Model studies of the stress-strain state of a plate with multilayer coatings under a contact distributed load on the contact area with normal forces changing according to a parabolic law have shown [11] that an increase in the strength and durability of the coating-base composition can be achieved by:

- application of hardening coatings with a high modulus of elasticity and a smooth gradient of change in properties in depth from the surface to the base (diffusion coatings);

- reducing the gradient of properties by depth, due to an increase in the coating thickness and an increase in the rigidity of the base;

- application of thin low-modulus films to the coating surface, providing an increase in the contact area and anti-friction properties.

These recommendations are clearly confirmed by the results of experimental studies (Table 3), in particular: application of a strengthening high-modulus coating of titanium and chromium nitride to a hard base; production of diffuse nitrided layers with a smooth gradient of change in hardness across the thickness; application of oxidizing films to nitrided layers; obtaining an optimal ratio of the hardness of the coating and the base during nitrohardening significantly increase the wear resistance and contact endurance of materials during rolling with slippage.

Table 3. Physical, mechanical and tribological characteristics and durability of samples after ion nitriding, heat treatment and other coatings during rolling friction tests in I-20 lubricant, ball load 150N (p0=3180 MPa)

No	Steel grade	Heat treatment / technology/ coating	Surface Micro-hardness, MPa	Base micro-hardness, MPa	Rolling track micro-hardness after testing, MPa	Coating thickness, μm	Wear intensity, $I \cdot 10^{-11}$	Pitting fatigue life, $N \cdot 10^{6}$ cycles
1	IIIX15	without heat treatment	3840	3340	3340	0	594	0,7
2	IIIX15	without heat treatment + ion nitriding	9180	2680	5400	300	312	1,08
3	IIIX15	without heat treatment + oxidation	6140	2680	5420	300	210	1,25
4	IIIX15	quenching	7210	7210	7130	0	8	25,1
5	IIIX15	quenching + ion nitriding	9180	4970	7160	300	8,4	24,2
6	IIIX15	quenching + ion nitriding	7660	5800	7300	290	7,5	26,4
7	IIIX15	quenching + ion nitriding	7860	5900	6350	120	13	22,8

8	ШХ15	quenching + TiN (CIB method)	10400	5120	5200	5	15	22
9	ШХ15	without heat treatment + TiN (CIB method)	14000	2680	3350	5	564	0,75
10	ШХ15	quenching + galvanic Cr coating	12000	7210	7140	5	7,6	28,4
11	ШХ15	quenching + galvanic Cr coating	12000	7210	7300	10	7	30,1
12	ШХ15	nitroquenching mode 1	8700	7200	7630	330	3,6	40,8
13	ШХ15	nitroquenching mode 2	7700	7420	7420	350	3,2	48,8
14	ШХ15	nitroquenching mode 3	7300	7200	7380	410	3,8	38,4
15	ШХ15	nitroquenching mode 4	8500	7410	7410	320	3,7	38,7

In the contact zone, under load, normal stresses arise with a maximum on the surface at the center of the contact area and shear stresses with a maximum at a certain depth. The presence of friction forces causes the maximum shear stress to shift from depth to depth.

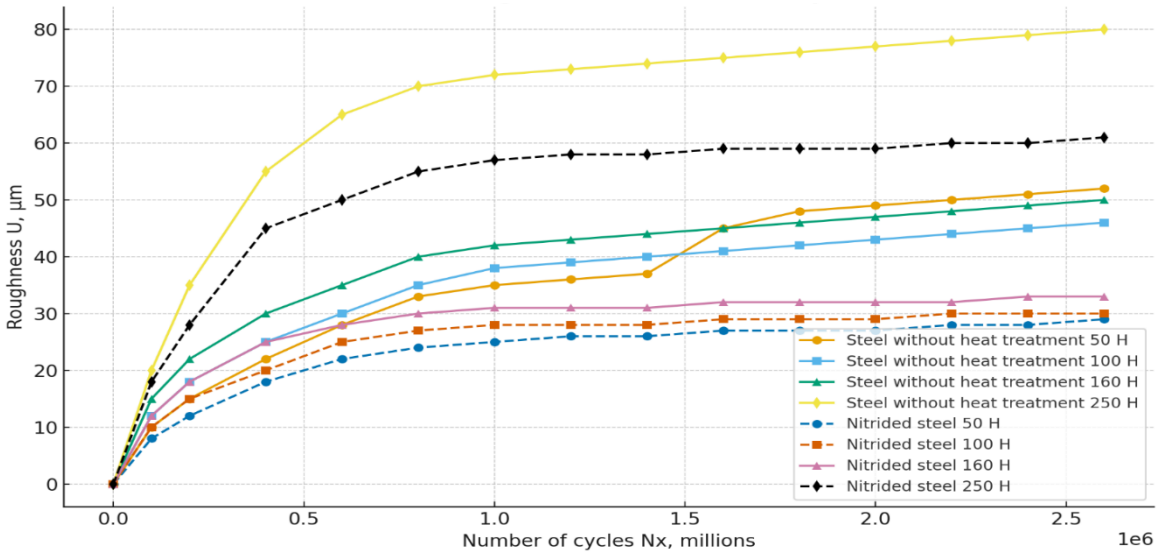


Fig. 3 Surface Roughness vs Number of Cicles

The graph 3 illustrates the dependence of the displacement amplitude (U) (µm) on the number of loading cycles (N \times 10^6) for steels with and without nitriding treatment under different applied loads (50, 100, 160, and 250 N). The solid curves correspond to steel without heat treatment, while the dashed ones represent nitrided steel (treated according to regime 1a). As the number of loading cycles increases, the displacement gradually grows for both materials, but the values for nitrided steel remain significantly lower throughout the entire range of cycles. This indicates a considerable increase in surface hardness and fatigue strength due to nitriding. For example, at the highest load of 250 N, the displacement of untreated steel reaches 70–80 µm,

while the nitrided surface exhibits only 35–40 μm deformation, which means a reduction of about 40–50%. The curves show that, during the initial stage of loading (up to 0.4×10^6 cycles), the deformation grows rapidly due to the running-in process, after which the growth rate stabilizes. Nitrided steel demonstrates an earlier stabilization of the curve, indicating the formation of a stable surface layer resistant to plastic deformation and microcrack propagation. In contrast, the untreated steel continues to accumulate deformation, showing signs of progressive surface fatigue. The overall tendency confirms that surface nitriding significantly improves the resistance of steel to cyclic contact loading by forming hard nitride phases ($Fe_{2-3}N$, Fe_4N) that reduce wear and prevent structural damage. Therefore, nitriding can be considered an effective method for enhancing the operational durability and dimensional stability of mechanical components such as gears, shafts, and cutting tools subjected to long-term cyclic stresses.

7. Conclusion

Thus, an analysis of operating conditions and the stress-strain state of gear teeth revealed that different areas of the tooth surface experience varying stress levels and types. The most hazardous surface areas include the root, the gullet, and the mid-tooth surface region located at the engagement pole. Therefore, it is clear that these surface areas require different surface layer properties. This can be achieved by strengthening the tooth surface, particularly in hazardous areas, by applying hardening coatings with a gradient structure across the depth; by creating optimal residual compressive stresses in the surface layers; and by strengthening the tooth core. This requires heat treatment of the tooth material to increase the core hardness and the application of hardening coatings using thermochemical treatment in hydrogen-free environments, with the physical and mechanical properties and phase composition of the coatings controlled depending on the operating conditions and stress-strain state of the gears.

REFERENCE LIST

[1]. Aleksandrov, V. I., & Sobolev, A. S. (1978). Materialy zubchatykh koles gruzovykh avtomobiley. Minsk: AN BSSR Institut problem nadezhnosti i dolgovechnosti mashin, 58 p.

[2]. Dymshits, I. I. (1960). Korobki peredach. Moscow: Mashgiz, 360 p.

[3]. Trubin, G. K. (1962). Kontaktnaya ustalost' materialov dlya zubchatykh koles. Moscow: Mashgiz, 404 p.

[4]. Kopf, I. A., Kornilov, V. V., & Efimov, E. V. (1998). Nestatsionarnaya termicheskaya model' zaedaniya i iznosa evolventnykh zubchatykh peredach. Tekhnika mashinostroeniya, 1(15), 54–59.

[5]. Zinchenko, V. M. (2001). Inzheneriya poverkhnosti zubchatykh koles metodami khimiko-termicheskoy obrabotki. Moscow: Izd-vo MGTU im. N. E. Baumana, 303 p.

[6]. Panasyuk, V. V., Andreykov, A. A., & Parton, V. Z. (1988). Mekhanika razrusheniya i prochnost' metallov: Spravochnoe posobie. Vol. 1: Osnovy mekhaniki razrusheniya. Kiev: Naukova Dumka, 488 p.

[7]. Kalachev, V. A. (1985). Vodorodnaya khrupkost' metallov. Moscow: Metallurgiya, 217 p.

[8]. Al'feld, G., & Fel'kly, I. (Eds.) (1981). Vodorod v metallakh. Vol. 2. Moscow: Mir, 430 p.

[9]. Karpenko, G.V., & Krip'yakevich, R.I. (1962). Vliyanie vodoroda na svoystva stali. Moscow: Metallurgizdat, 198 p.

[10]. Kaplun, V. G., & Pastukh, I. M. (2002). Enerho- i resursozberihayucha ekolohichno chysta tekhnolohiya ta obladnannya dlya zmitsnennya detaley mashyn. Mashynoznavstvo, 2, 49–51.

[11]. Kaplun, P. V. (1999). Doslidzhennya napruzheno-deformovanoho stanu plastyny z pokryttyamy pry kontaktnomu navantazhenni. Vymiryuval'na ta obchyslyuval'na tekhnika v tekhnolohichnykh protsesakh, 4, 179–182.

[12]. Kaplun, P. V. (2001). Kinetika iznosa staley s diffuzionnymi pokrytiyami pri kontaktnom tsiklicheskom nagruzhenii. Problemy tribolohii, 1, 119–124.

2025. Volume 9, Number 2 Pages 137-143

UDC: 621.311
DOI: https://doi.org/10.30546/09085.2025.02.328

ELECTRONIC TREATMENT OF COMPLEX STRUCTURES IN A TORCH VOLUMETRIC ELECTRIC DISCHARGE

Elchin GURBANOV[1*]

[1]*Baku Engineering University, Khirdalan, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-22
Received in revised form:2025-10-22
Accepted:2025-10-31
Available online

Keywords:
torch discharge;
corona discharge;
adhesion;
carbon-filled plastics;
wetting angle.
JEL classification: L60, L64,O33,C63

ABSTRACT

The influence of torch discharge treatment on carbon-filled plastics with epoxy binder and carbon-filled plastics with cloth layer on the surface was explored. Was elaborated the configuration of potential electrode for getting the stable torch discharge in positive half-period and corona in negative half-period of AC. The limiting wetting angle of materials surface after an activation by torch discharge was measured. Was shown an increase of adhesion between metal coating and treated surface. Was detected a big effect of electrical treatment in static regime when the electrical discharge influence on the surface is summed-up by the action of electron-ion components and active gaseous products action.

1. Introduction

Exposure to no equilibrium electrical discharges in gases can alter the physicochemical properties of the surface layer of materials, including increasing surface energy, which improves adhesion [1]. One interesting form is the torch discharge, which occurs in electronegative gases, including air, in non-uniform fields of a specific configuration [2] at interelectrode distances of 2-20 cm. It consists of a sequence of cathode-directed streamers and, in terms of its development, is intermediate between a corona discharge and a spark discharge. Like any transient form, the torch discharge is unstable, and its stabilization requires current limiting measures, for example, by using limiting resistors or creating a special field configuration.

In devices implementing a torch discharge [2,3], a metal "pin" serves as the anode, and a "plane" serves as the cathode. Conditions in the gap at a sufficient distance from the anode have little effect on the formation and stability of the torch discharge, as streamers are generated and formed in the region of high field strength near the anode. This characteristic creates the physical prerequisites for using torch discharge to modify the surface of products with a wide range of shapes, sizes, and electrical properties. For example, in [3], it was shown that an increase in adhesion properties after treatment in a torch discharge at a constant voltage is observed for wool fibers, fluoroplastic, and polyethylene threads.

[*]Corresponding author.

E-mail addresses: elqurbanov@beu.edu.az (Gurbanov Elchin Jalal).

At the same time, using torch discharge at a constant voltage to modify the surface of solid dielectric materials is difficult, even if a thin dielectric layer is located on a conductive substrate, since charge accumulation on the dielectric surface leads to "locking" of the discharge.

With alternating voltage, a torch discharge is ignited during each positive half-cycle, and an avalanche corona during each negative half-cycle of the applied voltage. The accumulated positive charge on the dielectric surface during the torch discharge is neutralized by the negative corona.

2. Methodology and experimental part

This study examined the effect of a torch discharge in air on the surface of composite materials-carbon fiber reinforced plastics with or without fiberglass sublayers. Experiments with electric discharge processing were conducted using the setup (fig. 1), where HVS is the high-voltage source; kV is a kilovoltmeter, Rlim and R are the limiting and measuring resistances; DG is the discharge gap; D_1 and D_2 are diodes and ELO is the electronic oscilloscope.

A refractory (molybdenum) "pin" electrode with a diameter of 1.5 mm and a cone-shaped end was used as the torch-forming device [4]. The cylindrical portion of the "pin" was covered with a dielectric nozzle. The working end of the "pin" faced the samples, which were positioned on a flat surface. The other end was attached directly to a limiting resistor (KEV-5) with a value of 10-20 Mom.

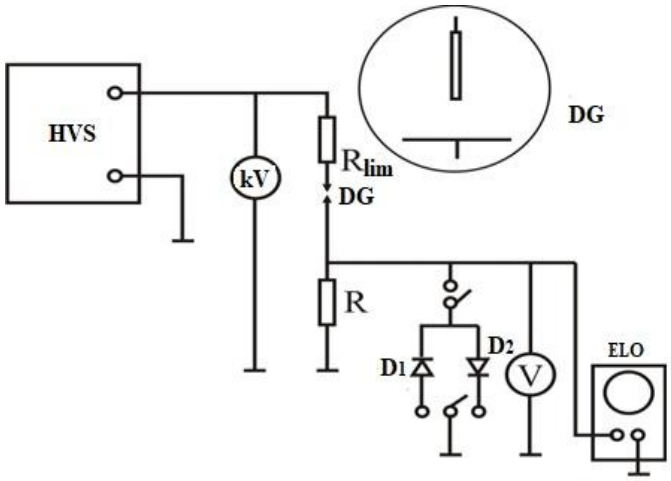


Fig. 1 Schematic diagram of the experimental setup for processing samples in a torch discharge

To determine the stable combustion mode of the torch discharge, geometric parameters characterizing the degree of influence of the dielectric nozzle on the field were varied: the distance to the work piece (2-10 cm) and the depth of the electrode insertion into the nozzle end (0-3 mm). As a result, an optimal anode design was experimentally determined that ensured a stable torch discharge during the positive half-cycle of alternating voltage at industrial frequency. A stable corona was ensured during the negative half-cycle. With a distance between the electrodes of L=4 cm, the torch wetting spot had a diameter of ~1.5 cm.

The presence of two diodes, D_1 and D_2, in the recording circuit allowed for separate measurement of the effective current values during each half-cycle of the applied voltage. Current during the negative half-cycle was recorded by connecting diode D_1, and current during the positive half-cycle by connecting diode D_2. The current-voltage characteristic of the torch-

forming device under alternating voltage is shown in figure 2, where curve 1 corresponds to the effective current value with the diodes disconnected, curve 2 to the positive half-cycle, and curve 3 to the negative half-cycle.

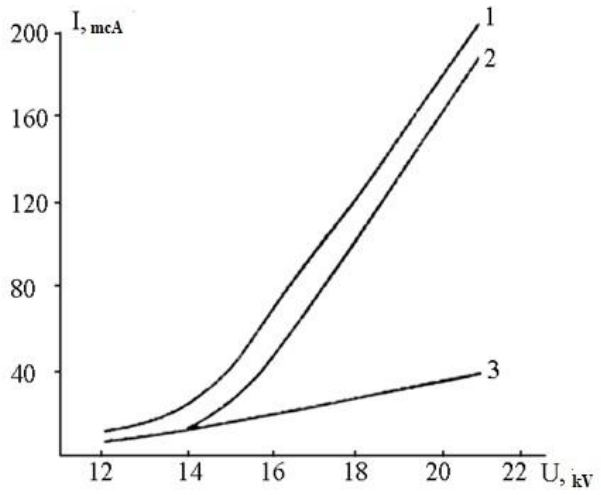


Fig. 2 Current-voltage characteristic of the torch-forming device at alternating voltage

The discrepancy between the total current and the sum of the currents during the positive and negative half-cycles of the alternating voltage is due to the imperfections of the diode characteristics. The appearance of a stable torch during the positive half-cycle is noted, according to oscillograms and visual observation, at U=12 kV. At U > 21 kV, spark channels are observed against the background of the stable torch. When a direct positive voltage of 22 kV is applied to the torch-forming device, a torch discharge is ignited. However, it has been noted that the torch discharge stability at direct voltage is significantly lower than with alternating voltage, and the stable torch currents for the same electrode configuration are lower. The stable torch range is 22-28 kV.

The applicability and effectiveness of torch discharge at direct and alternating voltages was tested when processing carbon fiber reinforced plastic (CFRP) plates and CFRPs with two fiberglass sublayers on each side. Since carbon fiber reinforced plastic (CFRP) is a conductive material, torch treatment can be used to modify its surface with both direct and alternating voltages. The surface of carbon fiber reinforced plastic (CFRP) plates with fiberglass sublayers is nonconductive, so torch treatment with alternating voltage is the only option.

The degree of modification was determined by the change in the wetting angle θ, which is related to the value of the work of adhesion forces by the Dupré-Young relationship:

$$W_A = \gamma(1 + \cos\theta) \quad (1)$$

where γ is the surface tension of the working fluid, θ is the wetting angle (for distilled water γ = 72.75*10-3 N/m or 72.75 mJ/m2).

The average value was calculated based on measurements at three points on the sample. The samples were treated at voltages of 15.5 and 19 kV with torch currents of 32 and 136 μA and negative half-cycle corona currents of 20 and 30 μA, respectively.

At 15.5 kV, the corona and torch currents differ by a factor of 1.5. This treatment of the carbon fiber reinforced plastic surface results in a monotonic increase in adhesion (curve 1, fig. 3). The main increase in adhesion is observed in the first 100 s of treatment, followed by insignificant

growth. After 5 min of treatment, adhesion $\cos\theta$=0.82. Treatment of the surface of carbon fiber reinforced plastics with a sublayer also leads to a monotonic increase (curve 2). The main increase is noted in the first 50 s of treatment. Five-minute treatment yields adhesion

$\cos\theta$=0.86. At a voltage of U=19 kV, the torch current exceeds the corona current by four times. In this case, a monotonic increase in adhesion is observed during treatment of both carbon fiber reinforced plastics and CFRPs with a sublayer (curves 3 and 4, respectively). After five minutes of treatment, adhesion $\cos\theta$=0.94 for carbon fiber reinforced plastics and $\cos\theta$=0.95 for carbon fiber reinforced plastics with a sublayer.

It was noted that with increasing discharge exposure time, the curves $\cos\theta$(t) reach saturation, which can be explained by the onset of a dynamic equilibrium between the processes of formation and destruction of groups that increase surface adhesion. Since, as the modified layer is removed, deeper layers of the sample will be exposed to the discharge.

When machining carbon fiber reinforced plastic (CFRP) samples with a sublayer, a characteristic peak appears on the dependence curve $\cos\theta$(t), which is reproduced under different machining conditions. This peak may be due to two different mechanisms increasing surface adhesion.

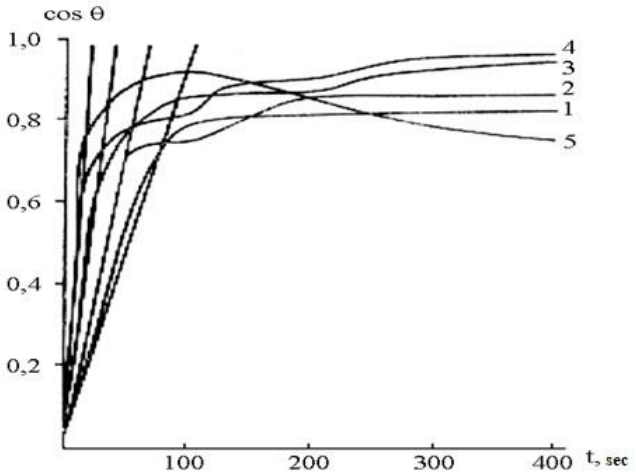


Fig. 3 Dependence of the change in the wetting angle of the surface of carbon fiber reinforced plastics (1, 3, 5) and CFRPs with a fiberglass sublayer (2, 4) on the duration of treatment in a torch discharge: 1, 2 – discharge current 90 μA at alternating voltage; 3, 4 – discharge current 150 μA at alternating voltage; 5 – discharge current 50 μA at constant voltage

Under the first machining condition (U=15.5 kV), the steepness of the initial portion of curve 3 is approximately three times greater than that of curve 1. This suggests that the increased machining effect is attributed to the effect of the torch discharge during the positive half-cycle. It is possible that some of the accelerated modification may also be attributed to the increase in negative corona current.

Under the second condition, at U=19 kV, higher values of adhesion work are achieved compared to the first condition, and the difference in wetting angles at different points on the same sample after machining is small, indicating uniform machining in this condition. Machining samples at U > 19 kV yielded higher average values, but due to the presence of sparks and their localization at certain points on the surface, machining of the samples was uneven. Furthermore, in this case, the appearance of spots on the surface of the carbon fiber reinforced plastics with a sublayer after processing was noted.

Under constant voltage, the carbon fiber reinforced plastic surface was treated at U=27 kV and a torch current of 50 μA. The dependence $\cos\theta$(t) for the treatment of the carbon fiber surface is shown in figure 3 (curve 5).

It can be seen that the dependence $\cos\theta$(t) peaks at 100 s ($\cos\theta$(t)=0.91) and then decreases, indicating that prolonged treatment degrades the previously achieved effect. Compared to treatment of carbon fiber reinforced plastic samples in a torch discharge under alternating voltage, the steepness of the $\cos\theta$(t) curve under constant voltage is greater. However, the ability to process at higher average currents and achieve better treatment results is an advantage of torch processing under alternating voltage. Furthermore, the use of alternating voltage simplifies the design and operation of the installation.

Experimental results show that a bipolar mode combining a torch discharge with a negative corona may prove optimal for practical use in electrical discharge surface modification of composite materials.

Furthermore, this mode enables surface modification of large dielectrics with equal torch and corona discharge currents, where the positive charge from the torch accumulated on the surface is compensated by the negative charge of the corona. In this case, a longer treatment period is required to activate the surface. However, for small samples, treatment with higher torch currents exceeding the corona current is possible, as this allows the charge to drain from the surface.

Since the effect of an air discharge on the surface of materials involves the action of the electron-ion component of the discharge and active oxygen-containing gaseous compounds (ozone, nitrogen oxides, and atomic oxygen), the objective was to determine which of these factors is decisive when treating the surface of carbon-fiber reinforced plastics and carbon-fiber reinforced plastics with a sublayer using a torch discharge under alternating voltage. To this end, the following experiments were conducted. First, during treatment near a flat electrode, the near-electrode zone was purged with an air flow under an excess pressure of 0.5 atm. This prevented oxygen-containing discharge products from coming into wetting with the sample surface, and only the electron-ion component of the discharge contributed to the surface modification. Treatment was then carried out while simultaneously using a metal mesh in wetting with the flat electrode. It was positioned above the sample surface parallel to their plane. In this case, the samples were exposed only to gaseous discharge products.

The wetting angle measurements after torch discharge treatment are presented in the table. The table lists the wetting angle values θ for carbon fiber reinforced plastics and carbon fiber reinforced plastics with a sublayer at various torch discharge treatment times. The first row lists the θ values for static treatment, the second row for blast treatment, and the third row for treatment with a mesh.

It is evident that the strongest treatment effect is achieved in static mode, i.e., with the simultaneous action of the electron-ion component of the discharge and gaseous discharge products. When exposed to the electron-ion component, the modification effect is slightly weaker. Gaseous discharge products have either a very weak effect or no effect at all.

As can be seen from the table, the effect of an electric discharge is determined by the combined action of the electron-ion component of the discharge and active gaseous compounds. A similar conclusion was reached in [5] when studying the electrical aging of polymer films in a barrier discharge.

Wetting angle measurement data for different types of carbon fiber reinforced plastics

Table 1.

Material	Material treatment time, min.		
Carbon fiber reinforced plastic	1	3	5
	Wetting angle θ		
	38	30	22
	41	33	29
	62	57	57
Carbon fiber reinforced plastic with sublayer	38	27	20
	40	35	34
	75	70	70

The role of gaseous discharge products in the modification of carbon fiber reinforced plastics with a sublayer is particularly evident. Since fiberglass is resistant to oxidation initiated by ozone and nitrogen oxides, as can be seen from the table, exposure to gaseous compounds alone does not cause changes in its adhesive properties. Therefore, if the effects of the discharge and its gaseous products are simply cumulative, then the absence of the latter should not affect the adhesion properties of carbon fiber reinforced plastics with a sublayer. However, it turned out that for these materials, the presence of ozone and nitrogen oxides in the discharge zone significantly enhances the change in adhesion properties.

In [6], it was shown that modification of carbon fibers in an ozone environment increases their surface adhesion as a result of oxidation. IR spectroscopy after ozonation of the fibers revealed the appearance of a broad band at 1050 cm-1, caused by the vibration of structural fragments containing single C-O-C bonds. It is likely that during electric discharge treatment, with the combined action of electron-ionic action and gaseous oxidants, adhesion should increase significantly more than that obtained in [6].

8. Discussion of results

To determine the qualitative change in the carbon fiber surface after treatment in a torch discharge, the IR transmittance spectra of the samples were studied. IR transmittance spectra were recorded for powder obtained by scraping from the carbon fiber surface to a depth of 10 μm. Powder, obtained from torch-exposed samples, as well as powder from untreated samples, were sintered into pellets using low-melting chalcogenide glass as the base material. IR spectra were then recorded using a SPECORD-751 setup.

Samples of fiberglass (the basis of the fiberglass sublayer in carbon fiber reinforced plastics with a sublayer) and cured epoxy resin were also pelletized and exposed to a flare discharge. The resulting IR transmittance spectra showed no noticeable changes in the treated fiberglass and epoxy resin samples compared to the untreated ones. This may be due to the fact that exposure to the discharge does not alter the chemical composition of the samples' surfaces.

In [7], no changes were noted in the epoxy resin after electric discharge treatment. In [8], non-stoichiometric oxygen was detected on the surface during glow discharge treatment. It is possible that such oxygen also forms after treatment of the fiberglass and epoxy resin, but IR spectroscopy cannot detect it. In CFRP samples treated in a torch discharge, an intense broad band appears at 1062 cm-1, which can be attributed to C-O vibrations. The observed increase in the work of surface adhesion forces in reducing the wetting angle is associated with oxidation of the CFRP surface and enrichment of the fiberglass surface with oxygen in the torch discharge.

Control and torch-treated CFRP samples and CFRPs with a sublayer were coated with an Al coating ~1 μm thick using thermal sputtering and magnetron methods. Coating adhesion to the surface, measured on a tensile testing machine, was 50-70% greater for the treated samples compared to the untreated ones.

When a metal coating is sprayed onto the sample surface after torch discharge treatment, metal atoms interact with non-stoichiometric oxygen and surface atomic groups, releasing corresponding oxidation energy at the film-substrate interface. This process is accompanied by the formation of valence bonds between substrate atoms and metal atoms through oxygen bridges.

9. Conclusion

A similar conclusion was reached when processing glass in a glow discharge [8], and the increase in adhesion of metals to glass after glow discharge treatment was explained by the formation of an intermediate layer of oxidized condensate. Better results were obtained when treatment in an oxygen-containing environment.

Treatment of carbon fiber reinforced plastics in a torch discharge in air leads to oxidation of the surface, accumulation of non-stoichiometric oxygen in it, which causes an increase in the adhesion of metal coatings applied to them.

REFERENCE LIST

[1]. Dzhuvarly Ch. M., Gorin Yu. V., Kurbanov E. D., Kulakhmetov F. Kh. (1987) Complex electric discharge surface treatment of materials. Reports of the 4th Scientific and Technical Conference "Vacuum Coatings - 87", Riga, 62

[2]. Bogdanova N. B., Pevchev B. G. (1975). Electrophysical properties of a torch discharge at constant voltage. J. Tech. Phys. 15(61) 61, 97-104

[3]. Andrianova R. L., Bogdanova N. B., Pevchev B. G. (1980). Torch discharge in some technological processes. Izvestiya AN SSSR. Power Engineering and Transport. 4. 102-108

[4]. Gasanov I.A., Gerasimov A.K., Gorin Yu.V., Dzhuvarly Ch.M., Dmitriev E.V., Kulakhmetov F.Kh., Gurbanov E.J. (1992). Device for implementing a torch discharge. USSR A.S., 1751826

[5]. Abramov R.Kh. (1986). Influence of gaseous products of electric discharge on molecular processes in polymer films. Abstract of Cand. Sci. (Tech.) Diss., Baku

[6]. Atyaksheva L.F. (1986). Study of the surface of carbon fiber treated with ozone. Surface,.3. 96-98

[7]. Bobrovskaya L.D., Burkhanova I.S., Petrenko T.N. (1982). Influence of a long-term applied electric field on the characteristics of an epoxy compound. Tr. All-Union Scientific Conf. "Breakdown and Electrical Aging", Baku

[8]. Kovalenko V.V. (1988). Effect of glow discharge plasma on the adhesion of metal condensates to silicon dioxide and silicon dioxide-based materials. Physics and Chemistry of Materials Processing. 1. 63-68

2025. Volume 9, Number 2 Pages 144-148

UDC: 621.574
DOI: https://doi.org/10.30546/09085.2025.02.346

EFFECT OF OIL ON THE HEAT TRANSFER OF MIXED REFRIGERANT BOILING IN EVAPORATOR TUBES

Akif BAKHSHIEV[1*], Asif GULIYEV[1], N. A. BAKHSHIEV[2]

[1]*Baku Engineering University, Khirdalan, Azerbaijan*

[2]*Ministry of Economy, Baku, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-24
Received in revised form:2025-10-29
Accepted:2025-11-14
Available online

Keywords:
heat transfer;
hydrodynamics;
refrigerant;
oil;
two-phase flow.

JEL classification: L64,Q41, Q42,O33

ABSTRACT

The paper presents the results of an experimental study of the effect of oil on the heat transfer rate at boiling of mixed refrigerant R406A. Since the air conditioning system is not a pure refrigerant, but a mixture of oil with a concentration of up to 8%, such an amount of oil affects both hydrodynamics and heat exchange in the evaporators. The experimental work covers the entire range of regime parameters typical for these systems. There is shown the process of changing oil concentration in the pipe, as the working fluid boils, proving that most of the oil pipe does not im- pair the heat exchange in the course of two-phase flow boiling. Different modes of refrigerant R406A boiling dynamics have been defined, and each mode is given a quantitative assessment in terms of the effects of the oil and explaining of this effect on the fluid flow and heat transfer based on visual observations and the experiment results. The main factor of the effect is the freon-oil foam, which increases the proportion of the wetted surface in the wave and stratified modes and the heat transfer rate to 30%.

1. Introduction

After the cessation of the use of popular freons, which were pure substances, in air conditioning systems due to environmental requirements, multicomponent refrigerants were proposed instead, for example R406A, which is a zeotropic mixture with significant non-isothermality during phase transitions. This refrigerant boils in the pipes of evaporators or air coolers. Evaporators with in-tube boiling have a lower charge of the working fluid, but also a lower heat transfer coefficient during boiling, which leads to the need to increase the heat exchange surface.

The boiling of the refrigerant in the pipe determines the complex hydrodynamics of the two-phase flow as the vapor content changes, which largely determines the intensity of heat transfer [1,3]. All this speaks to the ambiguity and complexity of the heat transfer process during in-tube boiling, which is aggravated by the presence of oil soluble in freon. At some operating concentrations, oil foams and distorts the hydrodynamics of the flow and the intensity of heat transfer during boiling. In addition, the presence of oil up to 8% significantly changes such properties of the working fluid as viscosity, thermal conductivity, which will also affect heat transfer.

* Corresponding author.
E-mail addresses: abaxsiyev@beu.edu.az (Bakhshiev Akif Bayram).

To lubricate compressor parts, synthetic oil BSE 32 is used, which is highly soluble with freon and circulates with it in the system. As confirmed by special studies [1], oil carryover from the compressor is 0.4÷1.2% of the working fluid and taking into account the separation of approximately 50% in the oil separator, carryover to the condenser and then to the evaporator will be 0.2÷0.6 %. During in-tube boiling in the evaporator, as the refrigerant moves, the oil concentration ξm increases. Since no liquid should enter the compressor, it boiling off in the evaporator should be almost complete, up to 90–95%. The remaining liquid refrigerant evaporates in the heat exchanger and suction pipe, and clean oil enters the compressor in small, non-hazardous portions.

2. Materials and results of the study

The experiments were carried out on a special stand with R406A refrigerant in a pipe 3.3 m long, 13 mm in diameter, with a wall thickness of 0.5 mm, made of 1X18HT steel. Range of changes in operating parameters: mass velocity $\omega\varrho$ = 30÷150 kg/(m2·s); boiling temperature t_0 = 5÷–20 °C; heat flux density q = 1÷10 kW/m2. Oil concentration at the pipe inlet ξm = 0÷4%. For visual observations, glass tubes are installed at the inlet and outlet of the pipe.

In Fig.1 shows the change in oil concentration along the length of the pipe in an evaporator with in-tube boiling.

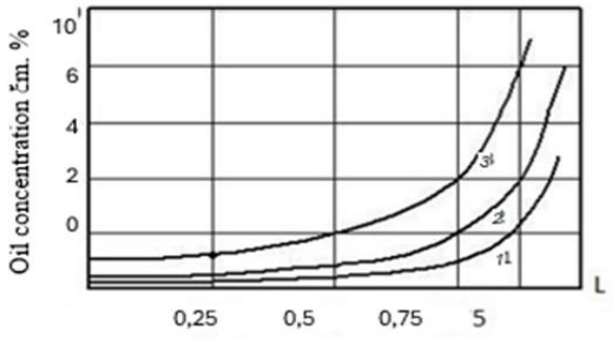


Fig.1. Dependence ξ_M = f(L) for complete boiling of the R406A refrigerant in an evaporator with in-tube boiling at the initial oil concen 1 – ξm = 0.25%; 2 – ξm = 0.5%; 3 – ξm = 1%

According to Fig.1 at the outlet of the evaporator pipe, the oil concentration ξm does not exceed 5–6%, and in most of its part ξm – 3%.

Previous studies have confirmed that when the oil concentration is less than 3%, the boiling intensity of the refrigerant becomes greater than when the pure working substance boils [2]. It is also noted that at ξm < 0.4% the hydrodynamics of the refrigerant flow will not change, and at high concentrations during boiling in the pipes foaming is observed, so the influence of oil on the hydrodynamics of boiling R406A in the evaporator will certainly be affected.

The degree of influence of oil on heat transfer depends on the flow mode. The refrigerant enters the evaporator after the throttling valve at a vapor content X = 0.1÷0.15 kg/kg. This corresponds to projectile or wave motion of the flow. The emulsion flow can be pumped.

In an emulsion flow, the addition of oil has virtually no effect on heat transfer, since heat transfer α is determined mainly by the speed of fluid movement, and at such an oil concentration ξm the properties of the working substance practically do not change and there is no foaming.

The heat transfer coefficient can be calculated using the formula

$$Nu = 0{,}021 \cdot \ Re^{0{,}8} \ \cdot Pr^{0{,}43} \ (1)$$

In the slug flow mode, the presence of oil also does not affect heat transfer, since the oil foam is located inside the bubble and does not come into contact with the heat transfer surface [4]. And in this mode, heat transfer αcon is determined by the speed of the flow, and boiling αboil intensifies heat transfer little.

When processing experimental data in slug mode, the dependence was obtained

$$\alpha = \alpha_{con} \sqrt[3]{1 + \alpha_{con} \cdot \alpha_{con}^3} \ (2)$$

In equation (2), the heat transfer coefficient during forced convection of liquid α_{con} is calculated according to (1) based on the true fluid velocity, and αboil - according to the dependence

$$Nu = 2{,}38 \cdot \ K_p^{0{,}25} \cdot (Pe \ \cdot K_t^{0{,}63} \cdot K_G^{0{,}5})^{0{,}75} \ (3)$$

Where $K_G = {}^{\tau}/_{q} \cdot \sqrt{{}^{\delta}/_{q} \cdot (p' - p'')}$ – characterizes the relationship between the heat of evaporation and free-free energy of the surface layer.

In the wave mode, the presence of oil foam significantly increases the wetted surface. Under conditions that corresponded to the wave regime for a pure refrigerant, in the presence of oil, the entire heat transfer surface turned out to be wetted by a wave or oil foam.

In Fig.2 shows a graph of the temperature distribution along the pipe wall in one section in relation to t_0 in the wave mode with and without oil.

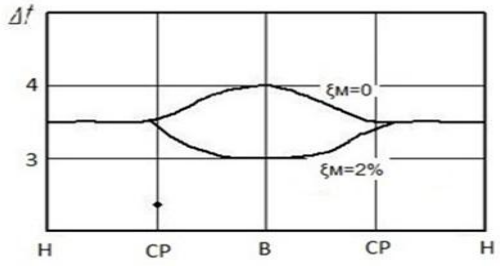


Fig.2. Temperature distribution along the pipe wall in wave mode with and without oil at ωQ = 100 kg/(m2 s); q = 2 kW/m2; P = 0.539 MPa

At ξ_M = 0, the temperature in the upper part of the pipe is higher than in the lower part, which is explained by the presence of a dry wall.

The intensity of heat transfer associated with boiling is not yet high, since in these regimes undeveloped boiling is observed.

Intensification of heat transfer in the presence of oil occurs mainly due to an increase in the proportion of the wetted surface in the upper part of the pipe. The effect of oil on α in this mode is ambiguous. As visual observations confirmed, at ξm < 0.4% this influence does not exist, since there is no foaming and the oil concentration practically does not change the properties of the working fluid. An increase in ξm > 3% leads to a decrease in heat transfer.

In Fig.3 and 4 show the increase in heat transfer coefficient in the presence of oil compared to α of pure R406A refrigerant.

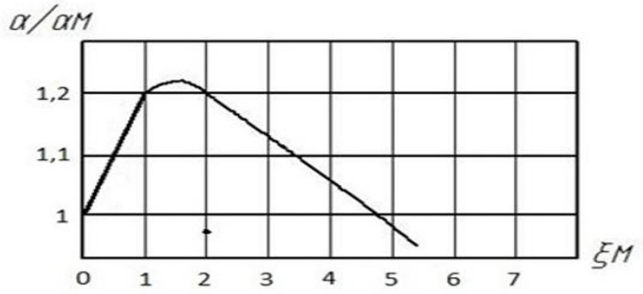


Fig.3. Ratio of coefficients α oil / α pure refrigerant in wave mode:
ωϱ = 50 kg/(m2 s); q = 2 kW/m2; t = –10 °C

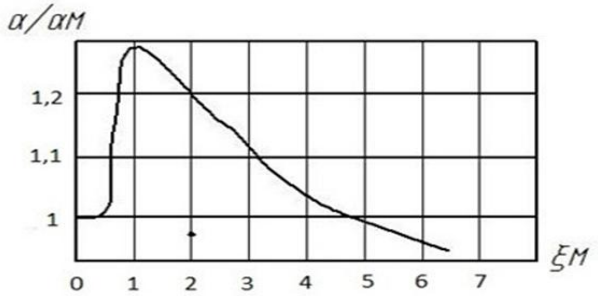


Fig.4. Ratio of coefficients α oil / α pure refrigerant in stratified mode:
ωϱ = 100 kg/(m2 s); q = 2 kW/m2; t = –20 °C

At the beginning of the regime, which corresponds to stratification during boiling of a pure refrigerant, the upper part of the pipe remains wetted with oil plugs, which move in large volumes along the surface of the liquid or fly in small portions in the vapor volume. In Fig.5 shows the temperature distribution in the pipe section at the parameters that determine the stratified flow regime of pure refrigerant.

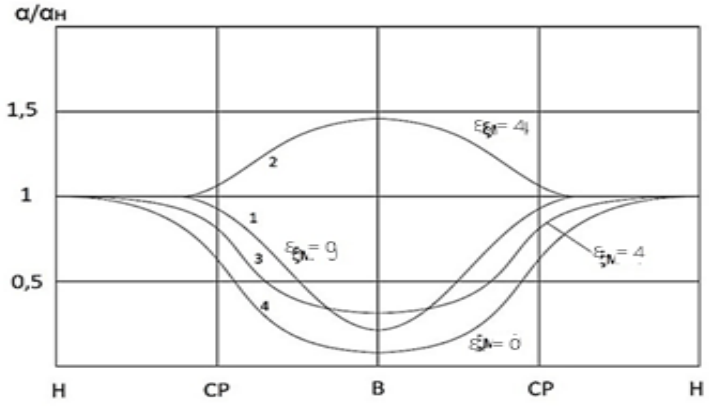


Fig.5. Ratio α/α_H in the pipe section in stratified mode: ωϱ = 50 kg/(m2 s); q = 2 kW/m2;
t = –10 °C; 1, 2 – at the beginning of the mode, 3, 4 – at the end of the pipe

The maximum heat transfer at the beginning of the mode is noted in the upper part, since it is wetted by foam. At the end of the pipe with a vapor content of X = 0.90÷0.95 kg/kg, the presence of a low-boiling stream at the lower generatrix of the pipe with a small cap of foam was visually noted. There is no foam in the steam area. Here, heat transfer determines the speed of steam movement and the boiling point of a freon-oil solution with a high oil content.

To determine the average heat transfer coefficient in a stratified mode of movement of a two-phase flow, W/(m2 K), we can propose the dependence

$$\alpha = \alpha_j \cdot \frac{F_w}{F} + \alpha_n \cdot (1 - \frac{F_w}{F}) \quad (4)$$

where, α_j is the heat transfer coefficient of the boiling liquid; αn is the heat transfer coefficient of moving steam; F_w – wetted surface of the pipe section, m²; F – pipe cross-sectional perimeter, m².

3. Conclusion

Calculation using formula (4) confirmed good agreement with the experimental results.

As a result of studying the effect of oil on the hydrodynamics and heat transfer of a two-phase boiling flow of refrigerant R406A:

– calculation formulas and criterion dependencies were obtained that make it possible to calculate the heat transfer coefficient during boiling of the R406A refrigerant with oil in the pipes of the evaporators of cooling systems;

– the influence of oil on the hydrodynamics of a two-phase flow was assessed;

– the influence of oil on heat transfer during boiling was determined both over the cross-section of the pipe and along its length.

REFERENCE LIST

[1]. Bukin V. G., Kuzmin A. Yu. Experimental study of small refrigeration units using a mixture of R22/142b // Refrigeration equipment. 1996. No. 5. pp. 12–15.

[2]. Komarov V.V., Shurshev V.F., Bukin V.G. Heat transfer during boiling of a mixture of R12/22 with oils on a pipe // Study of the thermophysical properties of substances and heat transfer processes in refrigeration technology. L.: Publishing house Leningrad. technol. Institute, 1989. pp. 90–94.

[3]. Bukin V.G., Kuzmin A.Yu., Ezhov A.V. The influence of adding oil during boiling of the R22/142b mixture inside a horizontal pipe on the hydrodynamics of the flow // Low-temperature and food technologies in the 21st century: materials of the International. scientific-technical conf. (St. Petersburg, June 06–07, 2001). St. Petersburg: Publishing house SPbGUNiPT, 2001. pp. 139–144

[4]. Bukin V. G., Kuzmin A. Yu., Ezhov A. V. Studies of heat transfer intensity and the mechanism of the boiling process of binary non-azeotropic mixtures // Improvement of energy systems and complexes: collection. scientific tr. Saratov: GTU Publishing House, 2000. pp. 24–32.

2025. Volume 9, Number 2 Pages 149-156

UDC: 620.197
DOI: https://doi.org/10.30546/09085.2025.02.352

MATHEMATICAL MODELING AND SIMULATION OF PROCESSES AFFECTING THE CORROSION KINETICS IN OIL PIPELINES ON THE ABSHERON PENINSULA

Nuriyya ABDULLAYEVA[1], Yegana ABDULAZIMOVA[2], Murad MAMMADZADA

[1]*Azerbaijan Technical University, Baku, Azerbaijan*

[2]*Baku Engineering University, Khirdalan, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-24
Received in revised form:2025-10-29
Accepted:2025-11-14
Available online

Keywords:
corrosion;
kinetics;
adsorption;
dislocation;
crystal lattice

JEL classification: L71, Q41, Q42, C63, C51, O33

ABSTRACT

The article presents the ***mathematical modeling and simulation of processes influencing the corrosion kinetics in oil pipelines on the Absheron Peninsula.*** *It has been determined that* ***a high concentration of chloride ions*** *and an* ***acidic environment*** *intensify the kinetics of corrosion. Under the influence of* ***radiation****, the main parameters of adsorption undergo significant changes. The* ***type of crystal lattice on the surface*** *directly affects the kinetics and energy of adsorption.* ***Dislocation processes*** *within crystal structures can alter the adsorption mechanism by modifying the surface characteristics of the material.* ***Dislocation lines*** *create local variations in energy and stress on the crystal surface, leading to preferential accumulation of adsorbates in these regions. The* ***stress fields*** *generated as a result of dislocation processes reduce the energy barrier for adsorption on the surface, thereby increasing the adsorption rate.* ***Simulation results*** *indicate that as the stress level rises, the degree of surface coverage by adsorbates also tends to increase. This finding demonstrates the strong influence of the* ***dislocation mechanism*** *on adsorption.* ***Thermodynamic parameters****, particularly* ***temperature*** *and* ***Gibbs free energy****, have a direct effect on adsorption and the kinetics of corrosion.*

*Corresponding author.

E-mail addresses: yabdulazimova@beu.edu.az (Abdulazimova Yegana Ayyub).

1. Introduction

Corrosion is a naturally occurring phenomenon that results in the gradual degradation of metals and alloys through chemical or electrochemical reactions with their environment [1]. This process leads to significant economic losses, reduced structural integrity, and compromised safety in various engineering systems [2]. Industrial infrastructures such as pipelines, bridges, ships, and power plants are particularly vulnerable to corrosion, making its control and prevention a major scientific and engineering challenge. The study of corrosion mechanisms provides a foundation for developing effective protective strategies, including coatings, inhibitors, and material selection [3]. Understanding the electrochemical behavior of metals in different environments allows for the optimization of corrosion-resistant materials and the design of long-lasting structures [4]. Moreover, recent advances in nanotechnology and surface

engineering have opened new perspectives for enhancing corrosion resistance through the development of smart and self-healing materials. Given the complexity of corrosion phenomena, research in this field integrates principles of electrochemistry, materials science, and environmental engineering [5]. Continuous investigation into the kinetics, thermodynamics, and environmental factors influencing corrosion contributes to improving sustainability and durability in modern industries [6,7].

2. Materials and methods

The analysis of **corrosion in oil pipelines on the Absheron Peninsula** is a highly relevant issue. The optimal solution to this problem lies in the **mathematical modeling and analysis of the kinetics of the corrosion process** in oil pipelines located on the Absheron Peninsula. The **simulation** was carried out considering **temperature, soil chemistry**, and other significant factors. Let us first consider the **exponential model of corrosion kinetics.** The kinetics of corrosion can be described by the following **exponential model:**

$$m(t) = m_0 \cdot \exp(-k \cdot t) \quad (1)$$

Here:

- m(t) – time-dependent mass (g/m²)
- m_0 – initial mass (7850 g/m²)
- k – corrosion rate constant (1.1×10^{-6} s^{-1})
- t – time (in seconds)

Figure 1 presents a graph illustrating the mass loss in oil pipelines due to corrosion over a 10-year period:

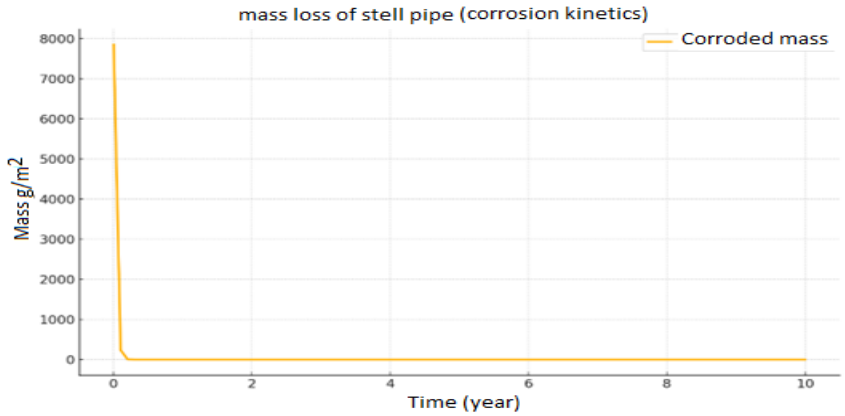


Figure 1. Graph illustrating the reduction of steel pipe mass in oil pipelines due to corrosion

As can be seen from the graph, the corrosion process leads to a very rapid mass loss, and within approximately one year, the pipe material may undergo complete degradation.

To determine the optimal solution to this problem, let us examine the influence of internal factors in the Absheron Peninsula. Specifically, we investigate the effects of internal parameters — chloride ions, pH, and microstructure — on the corrosion process in oil pipelines. The model has been calibrated using real data and expressed through a regression equation as follows:

Sample	Ph	[Cl^-] (ppm)	Microstructure (μ)	k_eff (1/s)
Sample 1	5.5	1000	0.9	2.0×10^{-6}
Sample 2	7.0	500	0.6	1.1×10^{-6}
Sample 3	8.0	200	0.3	5.0×10^{-7}

The effective corrosion constant is modeled as follows:

$$K\,eff\; =\; a_0\; +\; a_1\; \cdot\; [Cl^-]\; +\; a_2\; \cdot\; pH\; +\; a_3\; \cdot\; \mu \quad (2)$$

Here:

- $[Cl^-]$ – concentration of chloride ions (ppm)
- pH – acidity of the environment
- μ – microstructure parameter (defect density)

The following equation was obtained as a result of the multiple regression analysis:

$$K\,eff\; =\; 3.2 \times 10^{-6}\; +\; 1.5 \times 10^{-9} \cdot [Cl^-]\; -\; 2.8 \times 10^{-7} \cdot pH\; +\; 4.1 \times 10^{-7} \cdot \mu \quad (3)$$

The model indicates that high concentrations of chloride ions and an acidic environment accelerate corrosion kinetics.

Developing a corrosion model for oil pipelines in the Absheron Peninsula that accounts for external factors is one of the most important and significant aspects. In this study, the effects of external factors—temperature, humidity, soil acidity, and chloride concentration—on the corrosion process of oil pipelines in the Absheron Peninsula were mathematically modeled and simulated.

The corrosion rate is given by the following equation:

$$CR\; =\; k\; \cdot\; exp(-Ea\; /\; RT)\; \cdot\; f(RH, \rho, pH, Cl^-) \quad (4)$$

Here:

CR – corrosion rate (μm/h),

T – temperature (in Kelvin),

Ea – activation energy,

R - universal gas constant,

f – a function of humidity, soil resistivity, pH, and chloride ion concentration.

The effects of external factors influencing the corrosion of oil pipelines in the Absheron Peninsula were calibrated based on real samples, and the actual data used and the results obtained are presented as follows:

Sample №	Temperature (°C)	Humidity (%)	pH	Salinity (g/L)	Actual Corrosion (mm/year)	Predicted Corrosion (mm/year)
1.0	28.0	75.0	6.8	2.5	0.32	0.34
2.0	33.0	85.0	5.9	3.2	0.46	0.45
3.0	22.0	60.0	7.2	1.8	0.24	0.26
4.0	30.0	78.0	6.4	2.9	0.39	0.37
5.0	35.0	90.0	5.6	3.5	0.51	0.5

At the same time, the mathematical modeling and simulation of processes occurring in the nucleus—namely, corrosion and its primary controlling processes—are mainly governed by controlling factors. Specifically, nuclear processes, such as radioactive decay, nuclear reactions, and energy emission, can influence the adsorption of atoms or molecules on the surface. This influence primarily manifests in the following ways:

- **Thermal effect** – energy emitted from the nucleus increases the surface temperature, thereby altering the kinetics of adsorption.
- **Radiation-induced structural changes** – alters the atomic structure of the surface, which can either enhance or reduce adsorption.
- **Disturbance of electron clouds** – energetic particles can catalyze adsorption.

Nuclear processes affect surface corrosion and adsorption primarily through three mechanisms: energy activation, structural modification, and changes in electron levels.

In this section, the effect of nuclear energy (Q) on the adsorption transfer degree (θ) was evaluated by comparing real samples with the theoretical model. For calibration, a simplified Langmuir model was used, and the results were fitted using regression (Table 1).

Table 1. Effect of nuclear energy on the adsorption transfer degree

Q (MeV)	θ observed	θ model	θ regression
0.50	0.120	0.645	0.090
1.00	0.190	0.971	0.216
1.50	0.310	0.998	0.343
2.00	0.480	1.000	0.470
2.50	0.630	1.000	0.597
3.00	0.710	1.000	0.724

Mathematical modeling of the effect of radiation on the mechanisms of corrosion and adsorption is also a relevant issue. Radiation alters the adsorption mechanism, and these changes include:

- **Increased surface activity**;
- **Acceleration of desorption kinetics**;
- **Disturbance of electron balance due to ionization and polarization**.

Radiation affects the main parameters of adsorption. The model takes these changes into account, allowing prediction of both the kinetic and thermodynamic aspects of adsorption.

Let us consider the mathematical modeling of the effect of electron density on the mechanisms of corrosion and adsorption (on the Fe surface). In this study, the influence of electron density on the kinetics of adsorption and the degree of surface coverage on Fe is analyzed using a mathematical model.

Parameters for simulation conditions:

Parameter	**Value**	**Additional Note**
ϱ_0	1.5×10^{21} e/m^3	Initial electron density
Λ	0.1 μm^{-1}	Decay coefficient along the surface
Γ	0.01 s^{-1}	Temporal decay coefficient
k_a	0.03 s^{-1}	Adsorption rate constant
k_d	0.01 s^{-1}	Desorption rate constant
Δ	0.4	Electron density influence coefficient

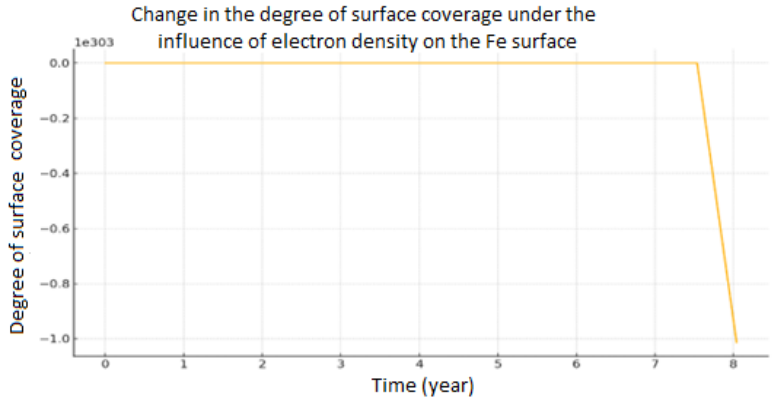


Figure 2: Graph of the change in surface coverage over time (on Fe surface)

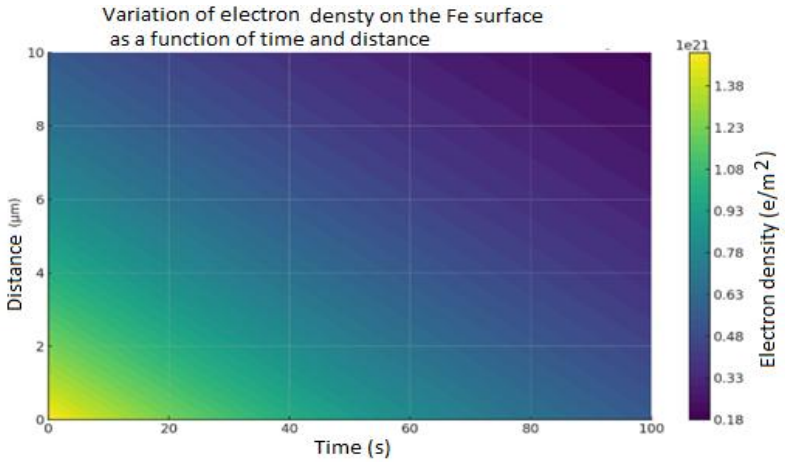


Figure 3: Graph of electron density variation with time and distance (on Fe surface)

Electron density directly affects the kinetics of adsorption. On the Fe surface, high electron density leads to a faster increase in surface coverage.

The kinetics of corrosion in oil pipelines are also influenced by the crystal lattice and dislocation mechanisms. The structure of crystal surfaces—namely FCC (Face-Centered Cubic), BCC (Body-Centered Cubic) and HCP (Hexagonal Close-Packed) significantly affects the adsorption mechanism. Since the atomic packing density, surface energy, and active area vary across these surfaces, the probability and kinetics of adsorbate molecules binding to the surface also change.

Let us present an extended Langmuir model that takes the type of crystal surface into account:

$$\theta(t) = (K\_i \cdot P / (1 + K\_i \cdot P)) \cdot (1 - e\^(-k\_i \cdot t)) \quad (5)$$

Here: - θ(t): Surface coverage (as a function of time)

- K_i: Adsorption constant (for crystal lattice type i)

- P: Gas pressure (or concentration)
- k_i: Adsorption kinetic constant

- i ∈ {FCC, BCC, HCP}

The adsorption constant is based on the Gibbs free energy:

$$K\_i = K\_0 \cdot e\^(-\Delta G\_ads, i / RT) = K\_0 \cdot e\^(-\gamma\_i \cdot A\_s, i / RT) \quad (6)$$

- γ_i: Surface energy of crystal lattice type i

- A_s,i: Surface area
- R: Universal gas constant
- T: Temperature (K)

Simulation conditions and parameters

Based on real data, the following results were obtained (for example, adsorption of N_2 or CO on Fe, Cr, or Ti surfaces):

Crystal Type	γ_i (J/m²)	A_s,i (nm²)	K_i (1/atm)	k_i (1/s)
FCC (Fe, Cu)	2.45	1.2	0.8	0.15
BCC (Cr, Mo)	2.15	1.6	0.6	0.12
HCP (Zn, Ti)	1.85	2.0	0.5	0.10
Crystal Type	γ_i (J/m²)	A_s,i (nm²)	K_i (1/atm)	k_i (1/s)

- **FCC metals** (e.g., Fe, Ni): High surface atomic density → high adsorption energies
- **BCC metals** (e.g., Cr): Moderate adsorption
- **HCP metals** (e.g., Ti, Zn): Weaker adsorption bonding

The type of crystal lattice on the surface directly affects the kinetics and energy of adsorption. The mathematical model, taking these differences into account, serves as a powerful tool for predicting the dynamics of the adsorption process. In crystal structures, dislocation processes influence the surface properties of the material, thereby altering the adsorption mechanism. Dislocation lines create local variations in energy and stress on the crystal surface, causing adsorbates to preferentially occupy these regions. In this study, the effect of dislocation mechanisms on adsorption was mathematically modeled and simulated.

Considering the dislocation mechanism, the kinetics of adsorption is modeled as follows:

$$\theta = [K \cdot P \cdot exp(-(Ea - \beta \cdot \sigma\_d)/(RT))] / [1 + K \cdot P \cdot exp(-(Ea - \beta \cdot \sigma\_d)/(RT))] \quad (7)$$

Here: θ - degree of adsorption coverage
K - adsorption constant
P - partial pressure
Ea - activation energy
σ_d - stress generated due to dislocations
β - coefficient representing the effect of stress energy
R - universal gas constant
T – temperature

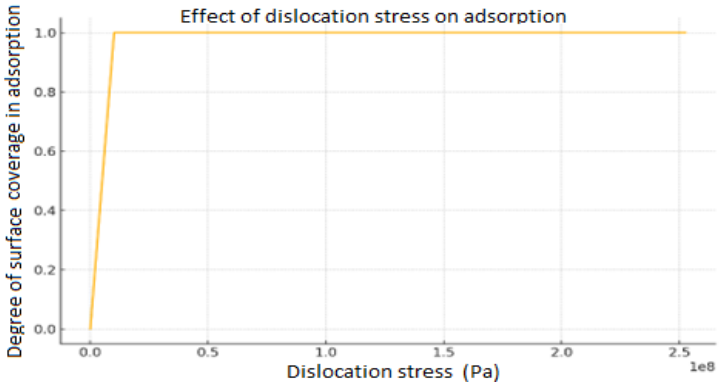


Figure 4. Effect of stress generated by the dislocation mechanism on adsorption coverage

Stress fields generated by dislocation processes reduce the energy barrier for adsorption on the surface, thereby increasing the adsorption rate. Simulation results showed that as the stress level increases, the surface coverage of adsorption also tends to increase. This finding demonstrates the dislocation mechanism has a strong influence on adsorption.

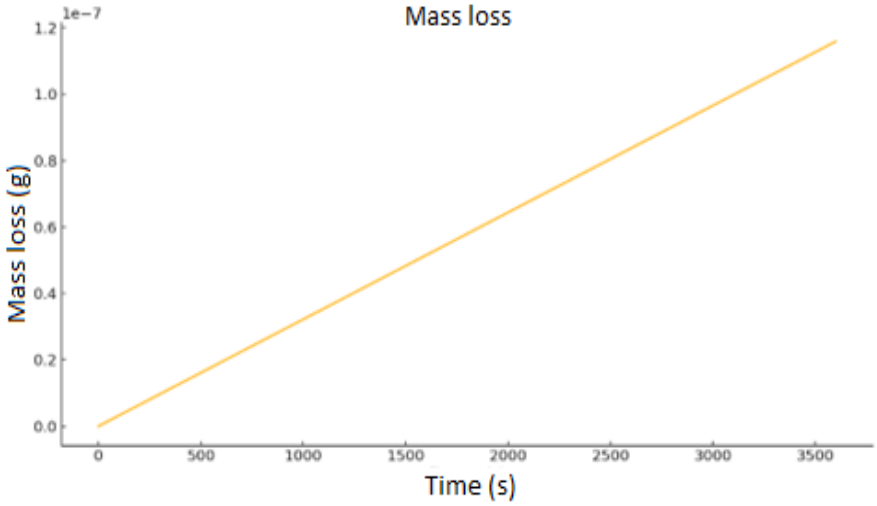


Figure 5. Effect of the dislocation mechanism on mass loss over time

An important aspect affecting the kinetics of corrosion in oil pipelines is the mathematical modeling and simulation of the thermal effect of adsorption. The thermodynamic basis and the thermal effect of adsorption in the Langmuir adsorption model are expressed as follows:

$$\theta = \frac{(K(T)\cdot C)}{(1 + K(T)\cdot C)} \quad (8)$$

Here: θ – surface coverage

C – concentration of the adsorbate,
K(T) – adsorption constant dependent on temperature

The temperature dependence of K(T) is expressed using the Van't Hoff equation:

$$K(T) = K_0 \cdot exp(-\Delta H\_ads / (RT)) \quad (9)$$

ΔH_ads – enthalpy of adsorption (kJ/mol),

R – universal gas constant

T – temperature (K).

In this case, the corrosion rate is modeled as follows:

$$V\_corr(T) = V_0 \cdot (1 - \theta(T)) = V_0 \cdot (1 - (K(T) \cdot C) / (1 + K(T) \cdot C)) \quad (10)$$

Here:

V_corr(T) – corrosion rate as a function of temperature

V_0 – corrosion rate in the absence of inhibitors

As the temperature increases, if ΔH_ads < 0 (exothermic adsorption), K(T) decreases, θ decreases, and the corrosion rate increases. The increase in desorption weakens the protective layer, exposing the surface to greater corrosion. Real data (for calibration):

Temperature (K)	Experimental Vcorr (mm/year)	Model Model (mm/year)
298	0.12	0.11
308	0.18	0.17
318	0.26	0.25
328	0.33	0.31
338	0.41	0.38

Taking the thermal effect of adsorption into account is critically important for inhibitor selection. Inhibitors with exothermic adsorption may lose their effectiveness at high temperatures. This model allows optimization of pipeline protection strategies according to thermo-mechanical conditions.

3. Conclusion

1. High concentrations of chloride ions and an acidic environment accelerate corrosion kinetics.
2. Radiation alters the main parameters of adsorption.
3. The type of crystal lattice on the surface directly affects the kinetics and energy of adsorption.
4. Dislocation processes in crystal structures influence surface properties of the material, thereby altering the adsorption mechanism. Dislocation lines create local variations in energy and stress on the crystal surface, causing adsorbates to preferentially occupy these regions.
5. Stress fields generated by dislocation processes reduce the energy barrier for adsorption on the surface, thereby increasing the adsorption rate. Simulation results showed that as the stress level increases, the surface coverage of adsorption also tends to increase. This finding demonstrates that the dislocation mechanism has a strong influence on adsorption.

REFERENCES

1. Fontana, M.G. (2021). Corrosion Engineering (3rd Edition). McGraw-Hill Book Company, New York
2. Revie, R.W., Uhlig H.H. (2020). Uhlig's Corrosion Handbook (3rd Edition). Wile, ISBN: 978-1-119-37941-1
3. Roberge, P.R. (2021). Handbook of Corrosion Engineering (2nd Edition)". McGraw-Hill Professional
4. Schweitzer, P.A. (2010). Fundamentals of Corrosion: Mechanisms, Causes and Preventive Methods. CRC Press, Boca Raton.
5. Yadav, M. et al. (2022). Corrosion inhibition performance of green inhibitors: A review". Journal of Molecular Liquids, DOI: https://doi.org/10.1016/j.molliq.2022.119924
6. Zhang, D. et al. (2023). Recent advances in corrosion-resistant coatings". Progress in Organic Coatings,. DOI: https://doi.org/10.1016/j.porgcoat.2023.107055
7. Li, X. et al. (2022). Modelling and simulation of corrosion under various environmental conditions". Corrosion Science, 2022. DOI: https://doi.org/10.1016/j.corsci.2022.110133

2025. Volume 9, Number 2 Pages 157-164

UDC: 621.311
DOI: https://doi.org/10.30546/09085.2025.02.110.9078

MODELING THE DEPENDENCE OF ARC OVERVOLTAGE ON GROUND FAULT RESISTANCE AND GROUND FAULT ANGLE

N.I. ORUJOV[1], H.B. GULIYEV[2], S.J. ALIMAMMADOVA[3]

[1]*Baku Engineering University, AZ0101, Hasan Aliyev str. 120, Khirdalan, Azerbaijan*

[2]*Azerbaijan Technical University, AZ1073, H. Javid avenue 25, Baku, Azerbaijan*

[3]*Azerbaijan State Oil and Industry University, AZ1010, Azadlig Avenue 16/21, Baku, Azerbaijan*

ARTICLE INFO

Article history
Received:2025-10-18
Received in revised form:2025-10-20
Accepted:2025-10-31
Available online

Keywords:
overvoltage multiplicity;
ground fault resistance;
ground fault angle;
regression equation;
correlation
JEL classification: L60, L64,O33,C63

ABSTRACT

In isolated neutral networks, the creation of artificial overvoltages is necessary for testing the insulation of electrical equipment under load conditions. In such scenarios, it becomes essential to establish the mathematical relationships between the parameters of single-phase non-stationary ground faults. During these faults, the dependencies between key parameters such as the multiplicity overvoltage, ground fault resistance and ground fault angle follow complex patterns. Therefore, it is important to develop adequate mathematical models that can describe the interdependencies of these parameters under practical conditions. This study addresses the analytical determination of the relationship between transient overvoltage, ground fault resistance, and fault angle in isolated neutral networks resulting from non-stationary ground faults. For this purpose, a regression equation was derived to express the dependence of the overvoltage multiplicity on ground fault resistance and fault angle, and the corresponding spatial representation was constructed. The obtained results confirm a strong correlation among these parameters, indicating their applicability for practical use.

[*]*Corresponding author.*

E-mail addresses: norucov@beu.edu.az (Orujov Najaf).

1. Introduction

It is well known that the failure of electrical equipment can lead to various injuries among operational personnel, disruption of technological processes, and severe accidents. Therefore, specific tests are conducted in advance to prevent such issues. In general, these tests are carried out in the following cases: commissioning of equipment or installations, after an accident, following scheduled or unscheduled maintenance, or after a certain period since the last inspection. Notably, high-voltage insulation testing of electrical equipment is mandatory for isolated neutral power networks with voltages up to 35 kV [1, 2].

In isolated neutral networks, performing high-voltage insulation tests under load conditions is of particular relevance for identifying potential equipment failures in advance and ensuring the uninterrupted power supply to consumers [3, 4]. Research conducted in this field indicates that various methods and tools have been proposed for such testing [5–7].

In [8], a method for testing insulation under load in isolated neutral networks is proposed. According to this method, artificial transient ground faults are created in the network based on Petersen's theory in order to test the insulation under load conditions. It should be noted that during such ground faults, the magnitude of overvoltage caused by arcing in the network

depends on the ground fault resistance, the ground fault angle, and the phase-to-ground capacitance of the network. Therefore, determining the ground fault resistance and angle in advance is essential for accurately selecting the test voltage. For this purpose, it is a relevant and important task to determine the dependence of arc overvoltage multiplicity—arising from transient ground faults in isolated neutral networks—on ground fault resistance, ground fault angle, and phase-to-ground capacitance of the network.

2. Problem statement

In general, to determine the dependencies among the aforementioned parameters, it is necessary to perform a numerical solution of the system of differential equations characterizing the transient process of single-phase earth fault in neutral isolated networks, using modern computational technologies. However, the numerical solution of this problem is considerably complicated due to the stiffness of the mentioned differential equations. In other words, since the system of differential equations is nonlinear, in some cases the stability of the solution is compromised and the results become distorted. Therefore, to overcome such difficulties, it is essential to obtain analytical expressions defining the dependencies of the arc overvoltage multiplicity (k) on the earth fault resistance (R_0), the earth fault angle (φ), and the phase-to-earth capacitance of the network (C_f). It should be noted that for this purpose, the analytical dependencies of the arc overvoltage multiplicity on the earth fault resistance have been studied in [9,10], on the earth fault angle in [11,12], on the phase-to-earth capacitance in [13,14], and on both the earth fault resistance and the phase-to-earth capacitance in [15]. Continuing these studies, this work considers the derivation of a regression model describing the dependence of single-phase arc overvoltage in neutral isolated networks on the earth fault resistance and earth fault angle.

3. Problem solving

The derivation of an analytical expression for the dependency of the arc overvoltage multiple, occurring during single-phase faults in an isolated neutral electrical network $(C_f = const)$, on the earth fault resistance and earth fault angle is considered. For this purpose, the results of experimental investigations conducted on a low-voltage model of an isolated neutral network $(C_f = 1mkF)$ are utilized, as presented in Table 1 [8].

Table 1.: *Dependence* $k = f(R_0, \varphi)$

R_0, Ohm	φ				
	30°	60°	90°	120°	150°
5	2,45	3,16	3,30	3,10	2,16
10	2,29	2,91	2,96	2,86	2,03
15	2,16	2,69	2,77	2,66	1,93
20	2,05	2,51	2,62	2,49	1,84
25	1,96	2,35	2,49	2,35	1,76
30	1,88	2,22	2,38	2,24	1,69

As seen from Table 1, the dependence between the multiple of the arc overvoltage, the ground fault resistance, and the ground fault angle can be approximated by the following regression equation [16]:

$$k = \frac{a}{R_0} + b\sin\varphi + c, \quad (1)$$

here, $a, b, c -$ are regression coefficients.

If we introduce the substitutions $\frac{1}{R_0} = x$ and $sin\,\varphi = y$ in equation (1), the regression equation can be written in the following form:

$$k = ax + by + c, \quad (2)$$

In other words, the dependence of the arc overvoltage multiplicity (k) on the conductivity of the ground fault circuit (x) and the sine of the ground fault angle (y) can be approximated by a linear regression equation (see Table 2). The choice of this model type is based on the assumption that if the variation of the output variable is directly proportional to the variations of the factor variables, then a linear model is considered adequate [17].

Table 2.: *Dependence* $k = f(x, y)$

x	y				
$[Sm]$	0,500	0,866	1,000	0,866	0,500
0,200	2,45	3,16	3,30	3,10	2,16
0,100	2,29	2,91	2,96	2,86	2,03
0,067	2,16	2,69	2,77	2,66	1,93
0,050	2,05	2,51	2,62	2,49	1,84
0,040	1,96	2,35	2,49	2,35	1,76
0,033	1,88	2,22	2,38	2,24	1,69

The regression coefficients of equation (2) are defined by the following well-known expressions [17]:

$$\left.\begin{aligned} a &= \frac{\sigma(k)}{\sigma(x)} \cdot \frac{r(k,x) - r(k,y)r(x,y)}{1 - r^2(x,y)}; \\ b &= \frac{\sigma(k)}{\sigma(y)} \cdot \frac{r(k,y) - r(k,x)r(x,y)}{1 - r^2(x,y)}; \\ c &= M(k) - aM(x) - bM(y). \end{aligned}\right\}, \quad (3)$$

where M(x), M(y) and M(k) denote the mathematical expectations (means) of the variables x, y, and k, respectively; σ(x), σ(y), and σ(k) denote the standard deviations of the variables x, y, and k, respectively; r(x,y) is the linear correlation coefficient between x and y; r(k,x) is the linear correlation coefficient between k and x; and r(k,y) is the linear correlation coefficient between k and y.

The numerical values of the necessary statistical indicators for determining the regression coefficients and identifying the target model are calculated based on the correlation matrix presented in Table 3. The following notation is used for the correlation matrix:

$$A_i = (x_i - M(x))^2;\ B_i = (y_i - M(y))^2;\ C_i = (k_i - M(k))^2;$$

$$E_i = (x_i - M(x)) \cdot (y_i - M(y));\ P_i = (k_i - M(k)) \cdot (x_i - M(x));\ T_i = (k_i - M(k)) \cdot (y_i - M(y));$$

$$L_i = \left| \frac{ax_i + by_i + c - k_i}{k_i} \right|;\ N_i = (ax_i + by_i + c - k_i)^2.$$

Table 3.: *Correlation Table*

i	x_i	y_i	k_i	A_i	B_i	C_i	E_i	P_i	T_i	L_i	N_i
1	0,200	0,500	2,45	0,01400278	0,06071797	0,001708	-0,02915854	0,0048911	-0,0101850	0,028818943	0,004985265
2	0,200	0,866	3,16	0,01400278	0,01430781	0,564502	0,01415447	0,0889078	0,0898709	0,022324562	0,004976684
3	0,200	1,000	3,30	0,01400278	0,06430781	0,794475	0,03000813	0,1054744	0,2260331	0,000706912	5,44199E-06
4	0,200	0,866	3,10	0,01400278	0,01430781	0,477942	0,01415447	0,0818078	0,0826940	0,003401812	0,00011121
5	0,200	0,500	2,16	0,01400278	0,06071797	0,061835	-0,02915854	-0,0294256	0,0612740	0,166947412	0,130036983
6	0,100	0,500	2,29	0,00033611	0,06071797	0,014082	-0,00451752	-0,0021756	0,0292407	0,08192638	0,035198042
7	0,100	0,866	2,91	0,00033611	0,01430781	0,251335	0,00219295	0,0091911	0,0599671	0,082049292	0,05700798
8	0,100	1,000	2,96	0,00033611	0,06430781	0,303968	0,00464915	0,0101078	0,1398125	0,027213051	0,006488404
9	0,100	0,866	2,86	0,00033611	0,01430781	0,203702	0,00219295	0,0082744	0,0539863	0,066001202	0,035631636
10	0,100	0,500	2,03	0,00033611	0,06071797	0,143388	-0,00451752	-0,0069422	0,0933073	0,035659403	0,005240108
11	0,067	0,500	2,16	0,00022500	0,06071797	0,061835	0,00369615	0,0037300	0,0612740	0,091211737	0,038815837
12	0,067	0,866	2,69	0,00022500	0,01430781	0,079148	-0,00179423	-0,0042200	0,0336518	0,058799026	0,025017553
13	0,067	1,000	2,77	0,00022500	0,06430781	0,130562	-0,00380385	-0,0054200	0,0916305	0,010814646	0,000897396
14	0,067	0,866	2,66	0,00022500	0,01430781	0,063168	-0,00179423	-0,0037700	0,0300633	0,048183977	0,01642739
15	0,067	0,500	1,93	0,00022500	0,06071797	0,229122	0,00369615	0,0071800	0,1179483	0,017089455	0,001087855
16	0,050	0,500	2,05	0,00100278	0,06071797	0,128642	0,00780299	0,0113578	0,0883791	0,076448938	0,024561259
17	0,050	0,866	2,51	0,00100278	0,01430781	0,010268	-0,00378782	-0,0032089	0,0121210	0,019072649	0,002291762
18	0,050	1,000	2,62	0,00100278	0,06430781	0,044662	-0,00803034	-0,0066922	0,0535920	0,019213916	0,002534162
19	0,050	0,866	2,49	0,00100278	0,01430781	0,006615	-0,00378782	-0,0025756	0,0097287	0,011193715	0,000776868
20	0,050	0,500	1,84	0,00100278	0,06071797	0,323382	0,00780299	0,0180078	0,1401252	0,028956347	0,002838724
21	0,040	0,500	1,96	0,00173611	0,06071797	0,201302	0,01026709	0,0186944	0,1105560	0,055378625	0,011781388
22	0,040	0,866	2,35	0,00173611	0,01430781	0,003442	-0,00498397	0,0024444	-0,0070174	0,029917391	0,004942915
23	0,040	1,000	2,49	0,00173611	0,06430781	0,006615	-0,01056624	-0,0033889	0,0206253	0,055629991	0,019187424
24	0,040	0,866	2,35	0,00173611	0,01430781	0,003442	-0,00498397	0,0024444	-0,0070174	0,029917391	0,004942915
25	0,040	0,500	1,76	0,00173611	0,06071797	0,420768	0,01026709	0,0270278	0,1598381	0,051964713	0,008364547
26	0,033	0,500	1,88	0,00233611	0,06071797	0,279488	0,01190982	0,0255522	0,1302688	0,03001239	0,003183588
27	0,033	0,866	2,22	0,00233611	0,01430781	0,035595	-0,00578140	0,0091189	-0,0225674	0,077668775	0,02973027
28	0,033	1,000	2,38	0,00233611	0,06430781	0,000822	-0,01225684	0,0013856	-0,0072696	0,092704827	0,048680901
29	0,033	0,866	2,24	0,00233611	0,01430781	0,028448	-0,00578140	0,0081522	-0,0201751	0,068046732	0,023233283
30	0,033	0,500	1,69	0,00233611	0,06071797	0,516482	0,01190982	0,0347356	0,1770868	0,079039472	0,017842737
⚜	**2,450**	**22,392**	**72,26**	**0,09819444**	**1,28615612**	**5,390747**	**0,00000000**	**0,4106667**	**1,998843**	**1,4663**	**0,5668**

The data array consists of a volume of n=30, and the values of the statistical indicators necessary for determining the coefficients, obtained as a result of calculations, are presented below.

$$\sum_{i=1}^{n} x_i = 2{,}45;\ \sum_{i=1}^{n} y_i = 22{,}392;\ \sum_{i=1}^{n} k_i = 72{,}26;$$

$$M(x) = \frac{\sum_{i=1}^{n} x_i}{n} = 0{,}082;\ M(y) = \frac{\sum_{i=1}^{n} y_i}{n} = 0{,}746;\ M(k) = \frac{\sum_{i=1}^{n} k_i}{n} = 2{,}409;$$

$$\sum_{i=1}^{n} (x_i - M(x))^2 = 0{,}09819444;\ \sum_{i=1}^{n} (y_i - M(y))^2 = 1{,}28615612;\ \sum_{i=1}^{n} (k_i - M(k))^2 = 5{,}390747;$$

$$\sum_{i=1}^{n} (x_i - M(x))(y_i - M(y)) = 0;\ \sum_{i=1}^{n} (k_i - M(k))(x_i - M(x)) = 0{,}4106667;$$

$$\sum_{i=1}^{n} (k_i - M(k))(y_i - M(y)) = 1{,}998843;$$

$$\sum_{i=1}^{n} \left| \frac{ax_i + by_i + c - k_i}{k_i} \right| = 1{,}4663;$$

$$\sum_{i=1}^{n} (ax_i + by_i + c - k_i)^2 = 0{,}5668.$$

Based on Table 3, variances, root mean square deviations, two-dimensional covariance coefficients (correlation moments), and two-dimensional correlation coefficients for individual quantities are calculated using well-known formulas. The obtained numerical values are as follows:

$$D(x)=\frac{\sum_{i=1}^{n}(x_i-M(x))^2}{n}=0{,}00327;\ D(y)=\frac{\sum_{i=1}^{n}(y_i-M(y))^2}{n}=0{,}04287;$$

$$D(k)=\frac{\sum_{i=1}^{n}(k_i-M(k))^2}{n}=0{,}17969;$$

$$\sigma(x)=\sqrt{D(x)}=0{,}0572;\ \sigma(y)=\sqrt{D(y)}=0{,}2071;\ \sigma(k)=\sqrt{D(k)}=0{,}4239;$$

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}(x_i-M(x))(y_i-M(y))}{n}=0;\ \operatorname{cov}(k,x)=\frac{\sum_{i=1}^{n}(k_i-M(k))(x_i-M(x))}{n}=0{,}01369;$$

$$\operatorname{cov}(k,y)=\frac{\sum_{i=1}^{n}(k_i-M(k))(y_i-M(y))}{n}=0{,}06663;$$

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{\sqrt{D(x)D(y)}}=0;\ r(k,x)=\frac{\operatorname{cov}(k,x)}{\sqrt{D(k)D(x)}}=0{,}56444;\ r(k,y)=\frac{\operatorname{cov}(k,y)}{\sqrt{D(k)D(y)}}=0{,}75911.$$

Then, based on expressions (3), the estimated values of the regression coefficients of equation (1) or (2) are obtained as follows:

$$a=4{,}18;\ b=1{,}55;\ c=0{,}91.$$

Thus, once the regression coefficients have been determined, the relationship (2) between the multiplicity of the arc overvoltage arising during single-phase non-stationary ground faults in isolated neutral networks, the conductivity of the ground fault loop, and the sine of the ground fault angle can be explicitly expressed as follows:

$$k=4{,}18x+1{,}55y+0{,}91,\ (4)$$

Let us verify the adequacy of the obtained regression dependence (4) between the conductivity of the single-phase transient ground fault circuit and the sine of the ground fault angle in relation to the overvoltage multiplicity during single-phase non-stationary ground faults. To do this, we can calculate the multiple correlation coefficient and assess its significance using the Fisher criterion [17].

The value of the multiple correlation coefficient is determined by the following well-known expression:

$$R=\sqrt{\frac{r^2(k,x)+r^2(k,y)-2r(k,x)r(k,y)r(x,y)}{1-r^2(x,y)}}=0{,}95.$$

A multiple correlation coefficient close to one ($R=0{,}95\to 1$) indicates that the relationship between the recurrence of arc overvoltages, the conductivity of the ground fault circuit, and the sine of the ground fault angle can be considered a strong linear correlation.

The significance of the multiple correlation coefficient is tested using the F-Fisher criterion. It is known that, at a significance level of α, the regression equation is considered adequate if the condition $F>F(\alpha,q_1,q_2)$ is satisfied [17], where q_1 and q_2 are the degrees of freedom.

The empirical value of the F-Fisher criterion, based on the given data, is determined as follows:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}, \quad (5)$$

Here, n is the number of experiments, n = 30; m is the number of factors, m = 2. According to equation (5), the value of $F = 125{,}4$.

The critical value of the Fisher criterion F is obtained from the table depending on the significance level (α) and the degrees of freedom (q_1, q_2) [17]:

$$\alpha = 0{,}05;\ q_1 = m = 2;\ q_2 = n-m-1 = 30-2-1 = 27;\ F(\alpha, q_1, q_2) = 3{,}35.$$

Since $F = 125{,}4 > F(\alpha, q_1, q_2) = 3{,}35$, the multivariate correlation coefficient ($R = 0{,}95$) and the statistical significance of the regression equation are confirmed.

The mean relative error and the mean square error of the approximation are determined by the following known expressions, respectively:

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{ax_i + by_i + c - k_i}{k_i}\right| \cdot 100\% = 4{,}89\%;$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(ax_i + by_i + c - k_i\right)^2}{n}} = 0{,}137.$$

The determination coefficient of $R^2 = 0{,}95^2 = 0{,}9025$ indicates that 90,25% of the variation in the arc overvoltage multiplication factor (k) is primarily caused by changes in the ground fault circuit conductance (x) and the sine of the ground fault angle (y), while the remaining variation 9,75 % is attributed to other unaccounted factors.

Thus, the dependence of the arc overvoltage multiplication factor on the ground fault resistance and the ground fault angle in neutral isolated networks under non-stationary ground faults can be explicitly expressed as follows:

$$k = \frac{4{,}18}{R_0} + 1{,}55\sin\varphi + 0{,}91, \quad (6)$$

As can be seen, during single-phase transient ground faults, the regression model obtained in the form of equation (6) represents a simple and practically applicable relationship between the parameters.

Based on the regression equation derived using the OriginLab software [18], a 3D (spatial) representation of the dependency of the arc overvoltage multiplication factor on the ground fault resistance and the ground fault angle has been constructed (Figure 1).

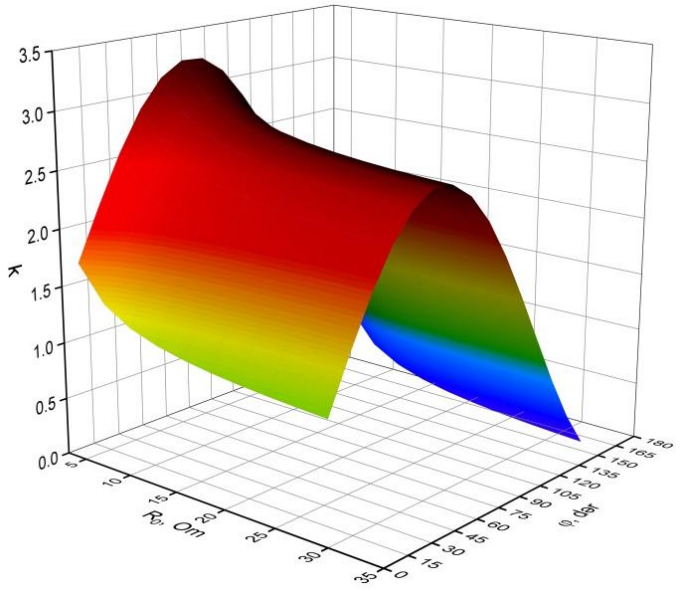


Figure 1. 3D visualization of the dependence of arc overvoltage multiplicity on ground fault resistance and ground fault angle

4. Conclusion

1. A practically implementable regression model has been developed to describe the dependence between the arc overvoltage multiplicity, ground fault resistance, and ground fault angle in neutral-isolated networks subjected to non-stationary ground faults governed by the Petersen theory.
2. The obtained regression model can be readily applied for insulation testing under load conditions in neutral-isolated networks of the Azerenergy system, as well as for the investigation and analysis of results related to non-stationary ground faults occurring in the network.

REFERENCES

[1]. Khazieva, R.T., Vasiliev, P.I., Aflyatunov, R.R. (2022). A study to establish the isolization of electrical equipment in the pierced area. *Electrical technology in passing*, 3(56), 65-69.

[2]. Patsch, R. (2018). Dielectric diagnostics of power transformers and cables – Return voltage measurements, theory and practical results. *VDE High Voltage Technology, ETG Symposium. https://www.vde-verlag.de/proceedingsde/454807130.html*

[3]. Shahmaev, I.Z., Gaisin, B.M., Shiryaev, O.V. (2016). A new method of taking management decisions at designing and developing electric power systems. *2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), IEEE, Chelyabinsk, Russia, 19-20 May*, 1-3.

[4]. Mudiraj, A.N. (2016) Improvement of Power Quality by mitigating harmonics in single phase AC distribution. *International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), IEEE, Pune, India, 09-10 September*, 83-88.

[5]. Антонов, А.И. (2017). Исследование уровня электромагнитных помех в сети 10/0,4 кВ с силовыми трансформаторами различной мощности при несимметричном характере нагрузки. Казан: Известия высших учебных заведений. *Проблемы энергетики*, № 9-10,. 65-76.

[6]. Rajasekhar, N.V., Babu, M.N. (2016). Harmonics reduction and power quality improvement by using DPFC. *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE, Chennai, India, 03-05 March*, 1754-1758.

[7]. Akdeniz, E., Bagriyanik, M. (2016). A knowledge-based decision support algorithm to reduce the impact of transmission system vulnerabilities. *International Journal of Electric Power and Energy Systems*. 78, 436-444.

[8]. Orucov, N.İ. (1998). Süni qeyri-stasionar yerlə qapanmalar əsasında neytralı izolə olunmuş şəbəkələrdə izolyasiyanın sınaq üsullarının işlənməsi və tədqiqi. *Texnika elmləri namizədi alimlik dərəcəsi almaq üçün təqdim edilən dissertasiya*. Bakı, 139 s.

[9]. Orucov, N.İ., Orucov, A.O. (2013). Qövs ifrat gərginliyin dəfəliyi ilə yerlə qapanma müqaviməti arasındakı asılılığın təyini, *Elmi əsərlər. AzTU*, *3*, 83-86.

[10]. Orucov, N.İ. (2019). Qövs ifrat gərginliyin yerlə qapanma müqavimətindən asılılığının reqressiya modeli. *"Azərbaycanın innovariv inkişafında mühəndisliyin rolu: hədəflər və perspektivlər" beynəlxalq elmi-praktiki konfransı, Bakı: BMU, - 29-30 noyabr,* 285-287.

[11]. Orucov, N.İ., Orucov, A.O. (2014). Qövs ifrat gərginliyi ilə yerlə qapanma bucağı arasındakı asılılığın təyini. *Energetikanın problemləri*, *2*, 36-39.

[12]. Orucov, N.İ., Mirili, T. (2021). Qövs ifrat gərginliyin yerlə qapanma bucağından asılılığının reqressiya modeli. *Gənc tədqiqatçıların V beynəlxalq elmi konfransı, - Bakı, Kitab 1. BMU, 29-30 aprel*, 113-116.

[13]. Orucov, N.İ. (2018). Qövs ifrat gərginliyin dəfəliyi ilə şəbəkənin yerə nəzərən faza tutumu arasındakı asılılığın təyini. *Azərbaycan Xalq Cümhuriyyətinin 100 illik yubileyinə həsr olunmuş "Təhsil - tədqiqat - istehsalat mexanizminin qurulması" mövzusunda Respublika Elmi-texniki konfransının materialları, Bakı: AzTU, - 4-5 aprel,* 254-257

[14]. Orucov, N.İ. (2021). Qövs ifrat gərginliyin şəbəkənin yerə nəzərən faza tutumundan asılılığının reqressiya modeli. *II beynəlxalq elm və texnologiya konfransı, - Bakı: BMU, - 26-27 noyabr*, 333-335.

[15]. Orujov, N.I., Guliyev, H.B., Alimammadova, S.J. (2024). Modeling of arc overvoltage dependence on ground circuit resistance and phase capacitance. *Reliability: Theory & Applications*, 19(4), 560-569.

[16]. Məmmədov, N.R., Məmmədov, B.M. (2005). *Eksperimentin nəticələrinin riyazi işlənməsi.* Bakı: Elm.

[17]. Orucov, E. (2018). *Ekonometrika.* Bakı: Elm. Origin(Pro), "Version 2022". OriginLab Corporation, Northampton, MA, USA.

INSTRUCTIONS FOR AUTHORS

1. "The Baku Engineering University Mechanical and Industrial engineering" accepts original unpublished articles and reviews in the research field of the author.
2. Articles are accepted in English.
3. File format should be compatible with **Microsoft Word** and must be sent to the electronic mail (**journal@beu.edu.az**) of the Journal. The submitted article should follow the following format:

- Article title, author's name and surname
- The name of workplace
- Mail address
- Abstract and key words

4. The title of the article should be in each of the three languages of the abstract and should be centred on the page and in bold capitals before each summary.
5. **The abstract** should be written in **9 point** type size, between **100** and **150** words. The abstract should be written in the language of the text and in two more languages given above. The abstracts of the article written in each of the three languages should correspond to one another. The keywords should be written in two more languages besides the language of the article and should be at least three words.
6. .**UDC** and **PACS** index should be used in the article.
7. The article must consist of the followings:

- Introduction
- Research method and research
- Discussion of research method and its results
- In case the reference is in Russian it must be given in the Latin alphabet with the original language shown in brackets.

8. **Figures, pictures, graphics and tables** must be of publishing quality and inside the text. Figures, pictures and graphics should be captioned underneath, tables should be captioned above.
9. **References** should be given in square brackets in the text and listed according to the order inside the text at the end of the article. In order to cite the same reference twice or more, the appropriate pages should be given while keeping the numerical order. For example: [7, p.15].

Information about each of the given references should be full, clear and accurate. The bibliographic description of the reference should be cited according to its type (monograph, textbook, scientific research paper and etc.) While citing to scientific research articles, materials of symposiums, conferences and other popular scientific events, the name of the article, lecture or paper should be given.

Samples:

a) ***Article:*** Demukhamedova S.D., Aliyeva İ.N., Godjayev N.M.. *Spatial and electronic structure af monomerrik and dimeric conapeetes of carnosine üith zinc,* Journal of structural Chemistry, Vol.51, No.5, p.824-832, 2010

b) ***Book:*** Christie ohn Geankoplis. *Transport Processes and Separation Process Principles.* Fourth Edition, Prentice Hall, p.386-398, 2002

c) ***Conference paper:*** Sadychov F.S., Aydın C., Ahmedov A.İ.. Appligation of Information – Commu-nication Technologies in Science and education. II International Conference."*Higher Twist Effects In Photon- Proton Collisions", Bakı, 01-03 Noyabr, 2007, ss 384-391*

References should be in 9-point type size.

10. The margins sizes of the page: - Top 2.8 cm. bottom 2.8 cm. left 2.5 cm, right 2.5 cm. The article main text should be written in Palatino Linotype 11 point type size single-spaced. Paragraph spacing should be 6 point.
11. The maximum number of pages for an article should not exceed 15 pages
12. The decision to publish a given article is made through the following procedures:

- The article is sent to at least to experts.
- The article is sent back to the author to make amendments upon the recommendations of referees.
- After author makes amendments upon the recommendations of referees the article can be sent for the publication by the Editorial Board of the journal.