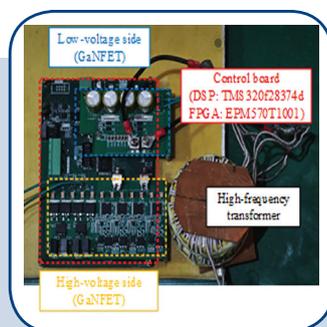
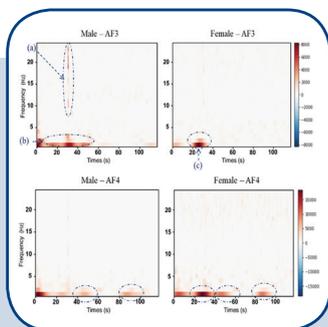
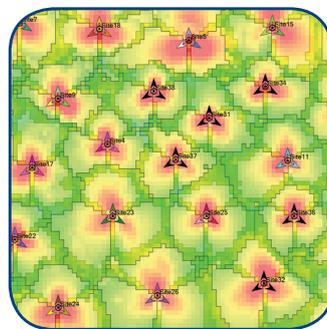
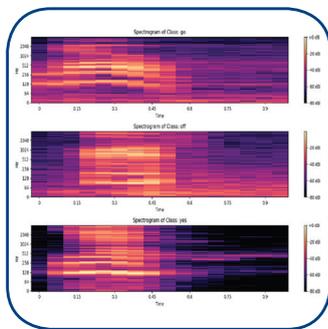


International Journal of Electrical and Computer Engineering Systems



INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia

Osijek, Croatia | Volume 16, Number 4, 2025 | Pages 265 - 346

The International Journal of Electrical and Computer Engineering Systems is published with the financial support
of the Ministry of Science and Education of the Republic of Croatia

CONTACT

**International Journal of Electrical
and Computer Engineering Systems
(IJECES)**

Faculty of Electrical Engineering, Computer
Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b, 31000 Osijek, Croatia
Phone: +38531224600, Fax: +38531224605
e-mail: ijeces@ferit.hr

Subscription Information

The annual subscription rate is 50€ for individuals,
25€ for students and 150€ for libraries.
Giro account: 2390001 - 1100016777,
Croatian Postal Bank

EDITOR-IN-CHIEF

Tomislav Matić
J.J. Strossmayer University of Osijek,
Croatia

Goran Martinović
J.J. Strossmayer University of Osijek,
Croatia

EXECUTIVE EDITOR

Mario Vranješ
J.J. Strossmayer University of Osijek, Croatia

ASSOCIATE EDITORS

Krešimir Fekete
J.J. Strossmayer University of Osijek, Croatia

Damir Filko
J.J. Strossmayer University of Osijek, Croatia

Davor Vinko
J.J. Strossmayer University of Osijek, Croatia

EDITORIAL BOARD

Marinko Barukčić
J.J. Strossmayer University of Osijek, Croatia

Tin Benšić
J.J. Strossmayer University of Osijek, Croatia

Matjaz Colnarič
University of Maribor, Slovenia

Aura Conci
Fluminense Federal University, Brazil

Bojan Čukić
University of North Carolina at Charlotte, USA

Radu Dobrin
Mälardalen University, Sweden

Irena Galić
J.J. Strossmayer University of Osijek, Croatia

Ratko Grbić
J.J. Strossmayer University of Osijek, Croatia

Krešimir Grgić
J.J. Strossmayer University of Osijek, Croatia

Marijan Herceg
J.J. Strossmayer University of Osijek, Croatia

Darko Huljenić
Ericsson Nikola Tesla, Croatia

Željko Hocenski
J.J. Strossmayer University of Osijek, Croatia

Gordan Ježić
University of Zagreb, Croatia

Ivan Kaštelan
University of Novi Sad, Serbia

Ivan Maršić
Rutgers, The State University of New Jersey, USA

Kruno Miličević
J.J. Strossmayer University of Osijek, Croatia

Gaurav Morghare
Oriental Institute of Science and Technology,
Bhopal, India

Srete Nikolovski
J.J. Strossmayer University of Osijek, Croatia

Davor Pavuna
Swiss Federal Institute of Technology Lausanne,
Switzerland

Marjan Popov
Delft University, Nizozemska

Sasikumar Punnekkat
Mälardalen University, Sweden

Chiara Ravasio
University of Bergamo, Italija

Snježana Rimac-Drlje
J.J. Strossmayer University of Osijek, Croatia

Krešimir Romić
J.J. Strossmayer University of Osijek, Croatia

Gregor Rozinaj
Slovak University of Technology, Slovakia

Imre Rudas
Budapest Tech, Hungary

Dragan Samardžija
Nokia Bell Labs, USA

Cristina Seceleanu
Mälardalen University, Sweden

Wei Siang Hoh
Universiti Malaysia Pahang, Malaysia

Marinko Stojkov
University of Slavonski Brod, Croatia

Kannadhasan Suriyan
Cheran College of Engineering, India

Zdenko Šimić
The Paul Scherrer Institute, Switzerland

Nikola Teslić
University of Novi Sad, Serbia

Jami Venkata Suman
GMR Institute of Technology, India

Domen Verber
University of Maribor, Slovenia

Denis Vranješ
J.J. Strossmayer University of Osijek, Croatia

Bruno Zorić
J.J. Strossmayer University of Osijek, Croatia

Drago Žagar
J.J. Strossmayer University of Osijek, Croatia

Matej Žnidarec
J.J. Strossmayer University of Osijek, Croatia

Proofreader

Ivanka Ferčec
J.J. Strossmayer University of Osijek, Croatia

Editing and technical assistance

Davor Vrandečić
J.J. Strossmayer University of Osijek, Croatia

Stephen Ward
J.J. Strossmayer University of Osijek, Croatia

Dražen Bajer
J.J. Strossmayer University of Osijek, Croatia

Journal is referred in:

- Scopus
- Web of Science Core Collection
(Emerging Sources Citation Index - ESCI)
- Google Scholar
- CiteFactor
- Genamics
- Hrčak
- Ulrichweb
- Reaxys
- Embase
- Engineering Village

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003
Published: quarterly
Circulation: 300

IJECES online
<https://ijeces.ferit.hr>

Copyright

Authors of the International Journal of Electrical
and Computer Engineering Systems must transfer
copyright to the publisher in written form.

TABLE OF CONTENTS

Sensorless Generalized Average Modeling-Based Control for the Resonant LC-DAB Converter	265
<i>Original Scientific Paper</i> Adelina Mukhametdinova Muhammad Mansoor Khan Ruifeng Zhang	
Stream-based Identification of Gender using Noninvasive Electroencephalographic Technology.....	279
<i>Original Scientific Paper</i> Bat-Erdene Gotov Tengis Tserendondog Uurtsaikh Luvsansambuu Munkhbayar Bat-Erdene Batmunkh Amar Dongsung Bae	
LTE Coverage Planning Based on Improved Grey Wolf Optimization	295
<i>Original Scientific Paper</i> Fekar Mohammed Riyadh El Mansour Guezouri Mustapha	
Enhancing Energy Efficiency in GAN-based HEVC Video Compression Using Knowledge Distillation.....	311
<i>Original Scientific Paper</i> Hajar Hardi Imade Fahd Eddine Fatani	
MLbFA: A Machine Learning-Based Face Anti-Spoofing Detection Framework under Replay Attack.....	321
<i>Original Scientific Paper</i> Vijay V. Chakole Swati R. Dixit	
A Hybrid Deep Learning Framework for Speech-to-Text Conversion as Part of Telemedicine System Integrated With 5G	333
<i>Review Paper</i> Medapati Venkata Manga Naga Sravan K Venkata Rao	
About this Journal	
IJECES Copyright Transfer Form	

Sensorless Generalized Average Modeling-Based Control for the Resonant LC-DAB Converter

Original Scientific Paper

Adelina Mukhametdinova

Shanghai Jiao Tong University, School of Electronics Information and Electrical Engineering
800 Dongchuan Rd, Shanghai, China
adelinam@sjtu.edu.cn

Muhammad Mansoor Khan*

Shanghai Jiao Tong University, School of Electronics Information and Electrical Engineering
800 Dongchuan Rd, Shanghai, China
mansoor@sjtu.edu.cn

Ruifeng Zhang

Shanghai Jiao Tong University, School of Electronics Information and Electrical Engineering
800 Dongchuan Rd, Shanghai, China
zrf333@sjtu.edu.cn

*Corresponding author

Abstract – The dual active bridge (DAB) converter is an efficient power conversion topology designed for applications that require bidirectional galvanic isolation and energy transfer. Among its various configurations, the resonant LC-DAB converter is notable for its ability to significantly reduce switching losses and enhance efficiency. While discrete-time control methodologies are widely employed for design and analysis of DAB converters, it is challenging to ensure performance stability during steady-state and transient operating modes. Furthermore, high-frequency and single point measurement of resonant LC-DAB possess challenge, as it may not reflect the behavior of resonant inductor current. To address these issues, a generalized average modeling-based approach is proposed, which minimizes output-side circulating current. This is achieved by aligning the output current with the secondary side converter's voltage. The proposed model eliminates the need for a current sensor and demonstrates low sensitivity during transient conditions. A two-stage control loop is utilized: an inner loop for current control and an outer loop with a PI controller for voltage control. The analysis and design procedure for the proposed control is detailed, followed by simulation and experimental results in order to demonstrate the effectiveness of the proposed method.

Keywords: Dual active bridge, resonant DC-DC converter, generalized average modelling, sensorless control

Received: October 25, 2024; Received in revised form: January 6, 2025; Accepted: January 19, 2025

1. INTRODUCTION

The DC-DC converter is essential as an interface between different DC-bus levels. It has been utilized in many applications, such as automotive applications [1], PV applications [2], DC transmission systems [3], and DC microgrids [4]. DAB DC-DC converter offers the advantages of a high step-down conversion ratio, high power transfer efficiency, bidirectional power flow, and galvanic isolation between the input and output stages [5-7]. The mode in which the converter operates with the current aligned with the secondary side volt-

age is preferred in many power electronics converters, as it minimizes switching losses [8]-[10]. In this mode, the current through the inductor returns to zero at the end of each switching cycle, but it does not stay at zero for a long period. One of the main challenges in power conversion systems is handling fluctuations in input voltage and varying load operations. These factors can lead to output voltage ripple and instability in the DAB system [11-13]. To solve this problem, a control loop can be used to ensure stability and maintain dynamic performance. However, measuring high-frequency current poses a challenge. Various authors in the existing

literature have proposed different methods for current measurement, including direct and indirect measurement of current, and current sensorless control.

The first category for current measurement uses a high-frequency current sensor to measure the current in real-time. This approach was employed in a robust digital nonlinear control system based on asymmetric half-cycle modulation [14]. It features model-based predictive current control with a compensator using a geometric sequence-control algorithm. This system demonstrates a smooth transient response, but its practical application is limited due to the requirement of a high-bandwidth current sensor. Another approach, a digital predictive controller for single-phase shift modulation, is proposed in [15-17]. The controller tracks the current reference within a single switching cycle to ensure that the transient DC offset is removed from the transformer current. It provides inherent over-current protection for the transformer but requires midpoint current sampling, leading to design complexity.

The second category involves indirect current measurement. In [18], the voltage across the extra input resistor is used for a feedforward controller. It measures the DAB converter's average input current under single phase shift modulation by analog integrator. Although this method shows a good dynamic response, however it requires additional complicated hardware. Another technique uses a load-current estimating method under a single-phase shift with switching-period delay compensation [19]. This method estimates the load current by measuring input and output voltage differences. The measurement noise is reduced by adopting the damping coefficient. However, this method increases the control complexity due to the load current estimating and input voltage sampling processes.

In the third category, current-sensorless control, eliminates the need for a current sensor. A model predictive control scheme under dual phase shift modulation

without a current sensor can predict the output power in the following switching cycle [20]. It uses the deviation between the reference voltage and input voltage. This method has low hardware cost and good dynamic performance. However, a sampling delay makes the feedback loop time constant large. Another paper proposes a current sensor-reduction control with an extended state observer under dual-phase-shift modulation [21]. The observed load current is directly used to predict phase-shift duty ratios after each switching cycle. This results in cost savings and increased reliability. Additionally, the proposed method can eliminate the steady-state error of the output voltage. However, the observer's accuracy depends on the precision of the model parameters.

Table 1 presents a comparative summary of various current measurement techniques, highlighting their respective advantages and limitations. Direct current measurement can be accurate but has high cost and limited practical application. In contrast, indirect current measurement tends to have a lower cost but may have implementation complexity. Sensorless current control offer elimination of sensors, which can reduce costs and simplify system design. Nevertheless, they often suffer from issues of accuracy.

Conventional converters typically utilize a pure inductor for damping purposes. These converters offer simplicity in analysis but have high switching and RMS currents, reducing efficiency [22, 23]. In contrast, the resonant DAB converter incorporates an inductor in series with a capacitor, forming a resonant tank that acts as the power impedance. This decreases the peak current and the transformer current's DC component, which enhances efficiency [24, 25]. However, its dynamic behavior can introduce significant stability challenges, necessitating the development of control strategies to mitigate oscillations and maintain performance across various operating conditions.

Table 1. Current measurement techniques comparison

Current measurement method	Controller type	Modulation technique	Benefit	Drawback	Reference
Direct measurement	Robust digital nonlinear control	Asymmetric half-cycle modulation	Smooth transient response	Limited practical application	[14]
Direct measurement	Digital predictive controller	Single phase shift	Inherent overcurrent protection ability	Midpoint current sampling requirement, high cost	[15]-[17]
Indirect current measurement	Feedforward controller	Single phase shift	Good dynamic response	Additional complicated hardware requirement	[18]
Indirect current measurement	Load-current estimating method	Single phase shift	Low measurement noise	Control complexity	[19]
Sensorless current control	Model predictive control	Dual phase shift	Low hardware cost and good dynamic performance	Large feedback loop time, inaccuracy	[20]
Sensorless current control	Sensor-reduction control with an extended state observer	Dual phase shift	Low-cost, good reliability, elimination of output voltage steady-state error	Accuracy dependence on the model parameters	[21]

Different works performed control strategies for resonant DAB converters. The paper [26] discusses phase

shift compensation with unified boundary control for resonant LC-DAB based on triple-phase-shift modula-

tion. It presents a steady-state analysis utilizing various modulation controls to minimize circulating current and enhance converter efficiency. However, it can work only at frequencies much higher than resonant frequency which may tend to higher switching losses. Another method uses frequency-domain analysis, representing current and voltage waveforms in a high-frequency link form [27]. This method accommodates variables related to multiple phase shifts. It can guarantee soft-switching operation, while minimizing the reactive power circulating between the bridges. However, it requires complex calculations to minimize reactive circulating power. The authors in [28] develop a small-signal model by extending the description function to predict the low-frequency behavior of the resonant LC-DAB. The model investigates the effects of switching frequency and phase shift as control parameters, incorporating a compensator into the control loop to enhance the dynamic response of resonant converters. However, accurately detecting both phase and gain relationships incurs high hardware costs and significant computational time.

Additionally, previous research primarily focused on adjusting the duty cycle, which becomes less effective when operating near the resonant frequency. These methods typically require the switching frequency to be at least 20% higher than the resonant frequency in order to function correctly [29]. Furthermore, it is observed that a significant portion of the existing literature has primarily employed discrete-time steady-state models, which can complicate the analysis and evaluation of a system's overall effectiveness. In contrast, using a continuous-time model facilitates a more straightforward evaluation of system behavior. The resonant LC-DAB model can be obtained through generalized average modeling (GAM) which provides continuous-time representations of AC state variables by utilizing complex Fourier coefficients [30, 31]. This approach provides a balanced trade-off between complexity and accuracy, enabling the ability of both zero current switching and zero voltage switching.

By considering the limitations discussed in prior research, the key contributions of this study, therefore are:

1. The GAM-based control has been proposed to effectively minimize circulating current on the output side by ensuring that the output current phase is aligned with the voltage of the secondary side converter.
2. By aligning the output current with the voltage, the model significantly reduces the current during switching, which enhances overall converter efficiency.
3. The model uses a sensorless control approach with continuous time representation. It allows for a more straightforward assessment of system behavior under varying conditions, eliminating the need for physical sensors that can introduce complexity and potential points of failure. The model has been verified through analysis and simulation.

The remainder of this paper is organized as follows. Section 2 presents the operation principle and GAM analysis of the resonant LC-DAB converter. Section 3 details the proposed control methodology, including inner and outer loop control. Section 4 presents the simulation and experimental results in detail with a comparison of the previous discussed method and proposed method. Finally, Section 5 concludes the paper.

2. STRUCTURE AND BASICS OF PROPOSED CONVERTER SYSTEM

2.1. CONVERTER TOPOLOGY

Fig. 1 shows a schematic of a resonant DAB converter. It consists of two full bridges on the primary and secondary sides, resonant inductance L , resonant capacitor C , and a high-frequency transformer characterized by a turn's ratio n . The full bridges consist of eight MOSFETs S_1 – S_8 with antiparallel diodes D_1 – D_8 . The terms C_i , C_o , R_o are the input, output capacitors, output resistance respectively. V_{in} , V_{out} , i_{out} , v_{AB} , v_{CD} , i_r , v_{LC} are input voltage, output voltage, output current, primary side voltage, secondary side voltage, inductor current, and LC-tank voltage, respectively.

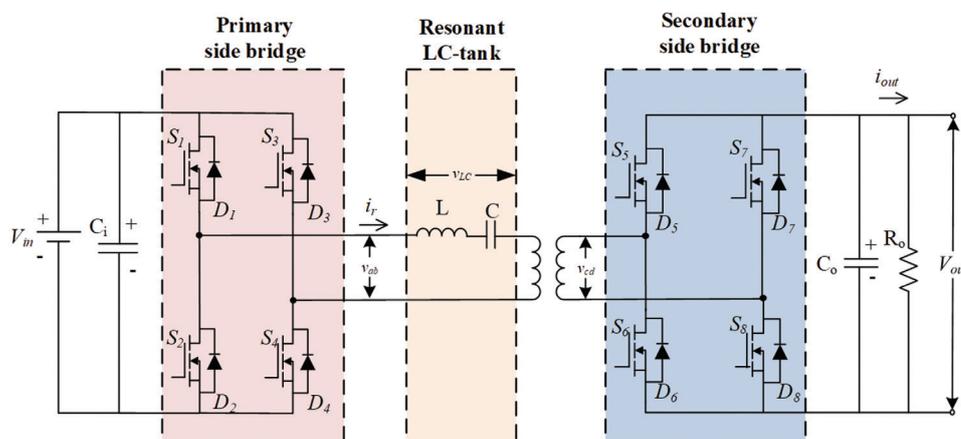


Fig. 1. Single-stage resonant LC-DAB

Fig. 2 illustrates the operating waveforms of resonant LC-DAB with output-aligned current. In stages 1 and 2, the inductor current is positive, conversely, in stages 3 and 4, the inductor current is negative. At time t_2 , the current approaches zero, allowing for ZCS to be achieved which leads to reduced switching losses. The normalized phase shift between leading and lagging leg of primary side is defined as d . The fundamental harmonic approximation is indicated by the dashed line. Notably, the phase shift between the resonant current and the first harmonic voltage of the secondary side is zero, while the phase shift between the resonant current and the first harmonic voltage of the primary side is represented as φ for illustration purpose. Both the duty cycle d and phase shift φ in a real implementation are smaller than those illustrated in the graph; however, they have been exaggerated for clearer understanding.

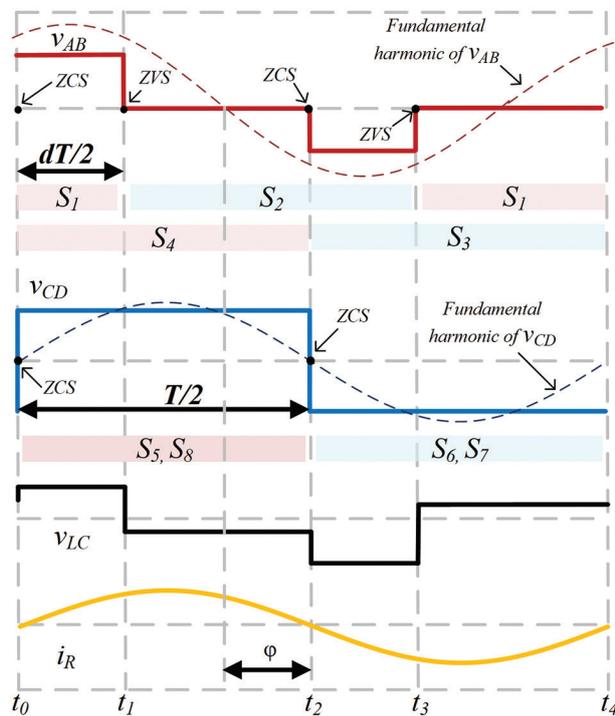


Fig. 2. Operating waveforms of the resonant LC-DAB

2.2. GENERALIZED AVERAGE MODELING

The GAM can formulate linear and time-invariant state equations. Throughout the analysis, τ represents the time referenced to the rising edge of s_1 . State variable $z(t)$ during the interval $t - T \leq \tau \leq t$ can be expressed using Fourier series expansion as follows [30]:

$$z(t) = \sum_{k=-\infty}^{+\infty} \langle z \rangle_k(t) e^{j\omega k t}, \quad (1)$$

Where $\omega = 2\pi f_s$, $\langle z \rangle_k$ - coefficient of the k th harmonics, which is defined by

$$\langle z \rangle_k(t) = \frac{1}{T} \int_{t-T}^t z(\tau) e^{-j\omega k \tau} d\tau. \quad (2)$$

The transformer current is entirely AC in a DAB converter, resulting in a significant ripple and a zero DC com-

ponent. To accurately represent the state variable averages in DAB converters, it is necessary to include multiple terms ($k = 0$ and $k = \pm 1$) in the Fourier expansion.

The product of 0th-order coefficient terms is

$$\langle mn \rangle_0 = \langle m \rangle_0 \langle n \rangle_0 + 2 \langle m \rangle_{1R} \langle n \rangle_{1R} + \langle m \rangle_{1I} \langle n \rangle_{1I}. \quad (3)$$

The 1st-order coefficient terms of two terms are

$$\langle mn \rangle_{1R} = \langle m \rangle_0 \langle n \rangle_{1R} + \langle m \rangle_{1R} \langle n \rangle_0 \quad (4)$$

$$\langle mn \rangle_{1I} = \langle m \rangle_0 \langle n \rangle_{1I} + \langle m \rangle_{1I} \langle n \rangle_0 \quad (5)$$

Where the lower indexes "R" and "I" represent the real and the imaginary parts of a complex number, respectively.

According to the graph in Fig. 2, the phase shift between the primary and secondary side legs is considered to be equal to zero, with the output voltage aligned with the leading leg of the secondary side. The 0th-order coefficients of $s_1(\tau)$ and $s_2(\tau)$ are zero due to a fixed 50% duty ratio. The switching function $s_1(\tau)$ of input side voltage has three states (1, 0, -1) and it can be defined as:

$$s_1(\tau) = \begin{cases} 1, & 0 \leq \tau < \frac{dT}{2} \\ 0, & \frac{dT}{2} \leq \tau < \frac{T}{2}, \frac{dT}{2} + \frac{T}{2} \leq \tau < T. \\ -1, & \frac{T}{2} \leq \tau < \frac{dT}{2} + \frac{T}{2} \end{cases} \quad (6)$$

Similarly, the output side switching function $s_2(\tau)$ has two states (1, -1) and can be represented as:

$$s_2(\tau) = \begin{cases} 1, & 0 \leq \tau < \frac{T}{2} \\ -1, & \frac{T}{2} \leq \tau < T \end{cases} \quad (7)$$

For output-aligned current, the 1st-order coefficients of primary and secondary side switching functions are obtained from the Fourier series expansion of the signal, derived in Eqs. (1)–(3), (6)–(7):

$$\left. \begin{aligned} \langle s_1(\tau) \rangle_{1R} &= \frac{\sin d\pi}{\pi} \\ \langle s_1(\tau) \rangle_{1I} &= \frac{-1 + \cos d\pi}{\pi} \\ \langle s_2(\tau) \rangle_{1R} &= 0 \\ \langle s_2(\tau) \rangle_{1I} &= -\frac{2}{\pi} \end{aligned} \right\}, \quad (8)$$

Using GAM and derived switching function (8), a large signal model for resonant DAB converter can be obtained [31]:

$$\frac{d}{dt} \langle i_s(\tau) \rangle_{1R} = \frac{\sin d_1 \pi}{\pi L_r} V_{in} - \frac{R_s}{L_r} \langle i_s(\tau) \rangle_{1R} - \frac{1}{L_r} \langle v_s(\tau) \rangle_{1R} + 2\pi f_s \langle i_s(\tau) \rangle_{1I}, \quad (9)$$

$$\frac{d}{dt} \langle i_s(\tau) \rangle_{1I} = \frac{-1 + \cos d_1 \pi}{\pi L_r} V_{in} - \frac{R_s}{L_r} \langle i_s(\tau) \rangle_{1I} - \frac{1}{L_r} \langle v_s(\tau) \rangle_{1I} + \frac{2}{\pi n L_r} \langle v_0(\tau) \rangle_0 - 2\pi f_s \langle i_s(\tau) \rangle_{1R}, \quad (10)$$

$$\frac{d}{d\tau}\langle v_0(\tau)\rangle_0 = -\frac{1}{C_0 R_0}\langle v_0(\tau)\rangle_0 - \frac{4}{n\pi C_0}\langle i_s(\tau)\rangle_{1I}, \quad (11)$$

$$\frac{d}{d\tau}\langle v_s(\tau)\rangle_{1R} = \frac{1}{C_r}\langle i_s(\tau)\rangle_{1R} + 2\pi f_s \langle v_s(\tau)\rangle_{1I}, \quad (12)$$

$$\frac{d}{d\tau}\langle v_s(\tau)\rangle_{1I} = \frac{1}{C_r}\langle i_s(\tau)\rangle_{1I} - 2\pi f_s \langle v_s(\tau)\rangle_{1R}, \quad (13)$$

Where $\langle i_s(\tau)\rangle_{1R}$, $\langle i_s(\tau)\rangle_{1I}$ are the 1st-order of real and imaginary parts of series inductance current; $\langle v_0(\tau)\rangle_0$ is the 0th-order of output voltage; $\langle v_s(\tau)\rangle_{1R}$, $\langle v_s(\tau)\rangle_{1I}$ are the 1st-order real and imaginary parts of resonant capacitor voltage.

The current on the output side can only have two states: 1) the i_{out} when transistors S6 and S7 are OFF; and 2) the $-i_{out}$ when transistors are S5 and S8 OFF. Therefore,

$$i_{out}(\tau) = s_2(\tau)i_s(\tau). \quad (14)$$

Only the 0th-order coefficient of output current is considered, as both are DC components. Therefore, by applying Fourier expansion in (3), the 0th-order coefficient of output current can be found:

$$\begin{aligned} \langle i_{out}(\tau)\rangle_0 &= \langle s_2(\tau)\rangle_0 \langle i_s(\tau)\rangle_0 \\ &+ 2\langle s_2(\tau)\rangle_{1R} \langle i_s(\tau)\rangle_{1I} \end{aligned} \quad (15)$$

3. PROPOSED CONTROL METHODOLOGY

3.1. OUTPUT-ALIGNED CONTROL METHOD

Existing research employed discrete-time modeling and steady-state analysis for zero voltage switching (ZVS) and zero current switching (ZCS) [17, 32]. However, the solutions derived were overly complex and not directly applicable to microcontrollers. To address this, the researchers made approximations that are only effective at high frequencies. Additionally, the stability and sensitivity of ZVS and ZCS during transients have not been guaranteed or thoroughly analyzed. While these methods may perform well under steady-state conditions, they do not necessarily ensure ZVS and ZCS during transient.

According to (7), the secondary side switching function is aligned with the leading leg, causing the fundamental component to be oriented along the imaginary axis. For optimal performance, the inductive current should mainly consist of an active component with a minimum reactive component, as it does not contribute to energy transfer. To achieve ZVS and ZCS, the inductive current should predominantly feature an imaginary component with a minimized real part.

The duty cycle d can be calculated differently depending on load conditions in the output-aligned current mode. Nevertheless, the proposed implementation has been approximated using the Taylor approximation [8]:

$$d = \frac{V_{out}}{V_{in}}. \quad (16)$$

I_{ref} is the desired output current value during steady-state for the proposed controller design. Therefore,

$$I_{ref} = -\frac{4}{\pi}\langle i_s(\tau)\rangle_{1I}. \quad (17)$$

Output-aligned current mode can be achieved using the following constraints:

1) output current is following I_{ref} ; 2) real part of inductance current is equal to zero $\langle i_s(\tau)\rangle_{1R} = 0$; and (3) imaginary part of voltage is close to zero $\langle v_s(\tau)\rangle_{1I} \approx 0$ (due to small series resistance of LC-tank).

Therefore, during steady-state, it should satisfy equations derived using Eqs. (9), (13), (17):

$$\langle i_s(\tau)\rangle_{1I} = -\frac{\pi}{4}I_{ref}, \quad (18)$$

$$\begin{aligned} \frac{\sin d_1 \pi}{\pi L_r} V_{in} - \frac{R_s}{L_r} \langle i_s(\tau)\rangle_{1R} - \frac{1}{L_r} \langle v_s(\tau)\rangle_{1R} \\ + 2\pi f_s \langle i_s(\tau)\rangle_{1I} = 0, \end{aligned} \quad (19)$$

$$\frac{1}{C_r} \langle i_s(\tau)\rangle_{1I} - 2\pi f_s \langle v_s(\tau)\rangle_{1R} = 0. \quad (20)$$

From these equations, the frequency can be found:

$$f_{s_{1,2}} = \frac{(-b \pm \sqrt{b^2 - 4ac})}{2a}, \quad (21)$$

with coefficients,

$$\left. \begin{aligned} a &= I_{ref} \pi^3 \\ b &= -\frac{2V_{in} \sin d_1 \pi}{L_r} \\ c &= -\frac{\pi I_{ref}}{4L_r C_r} \end{aligned} \right\} \quad (22)$$

3.2. SENSORLESS CURRENT CONTROL

The proposed current control diagram of the converter is illustrated in Fig. 3.

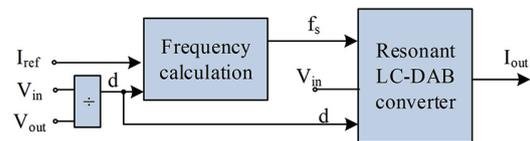


Fig. 3. Proposed current control diagram of the converter

The duty cycle is calculated by dividing the measured output voltage by the input voltage, as derived in (16). Switching frequency f_s is calculated by frequency calculation block, as described in (21), which is a function depending on the reference output current I_{ref} and the duty ratio d :

$$f_s = f(I_{ref}, d). \quad (23)$$

By processing these inputs, the resonant LC-DAB converter block, which utilizes the GAM calculation, determines output current, as described in (9)—(13):

$$X = f(w, d, V_{in}), \quad (24)$$

$$I_{out} = -\frac{4}{\pi} X(2), \quad (25)$$

Where,

$$[(\langle i_s(\tau) \rangle_{1R} \ \langle i_s(\tau) \rangle_{1I} \ \langle v_o(\tau) \rangle_0 \ \langle v_s(\tau) \rangle_{1R} \ \langle v_s(\tau) \rangle_{1I}]^T.$$

3.3. SMALL-SIGNAL MODEL

To design controllers and analyze stability for power converters, it is essential to obtain the small-signal control-to-output transfer function. This function shows the converter's dynamic response to small perturbations in the control signal.

The small signal model can be derived by perturbing the system in Eqs. (9)—(13) around steady-state values. The deviations of the state variables are expressed as follows:

$$\left. \begin{aligned} d &= D + \Delta d \\ i_{s1R} &= I_{s1R} + \Delta i_{s1R} \\ i_{s1I} &= i_{s1I} + \Delta i_{s1I} \\ v_o &= V_o + \Delta v_o \\ v_{s1R} &= v_{s1R} + \Delta v_{s1R} \\ v_{s1I} &= v_{s1I} + \Delta v_{s1I} \end{aligned} \right\} \quad (26)$$

The variables in capital letter X denote the steady-state values, the large-signal states are defined by the variables in lowercase letters x and the perturbations are represented by the Δx variables.

Steady-state values can be derived by solving:

$$\frac{d}{dt} [(\langle i_s(\tau) \rangle_{1R} \ \langle i_s(\tau) \rangle_{1I} \ \langle v_o(\tau) \rangle_0 \ \langle v_s(\tau) \rangle_{1R} \ \langle v_s(\tau) \rangle_{1I}]^T = 0. \quad (27)$$

Then, the small-signal model can be found by perturbation around steady-state value and have the form:

$$\Delta \dot{x} = A \Delta x + b \Delta I_{ref}, \quad (28)$$

$$\Delta \dot{v}_o = q \Delta x, \quad (29)$$

Where expressions of matrices are given in the Appendix A.1.

The control-to-output transfer function can be obtained using:

$$G_s(s) = \frac{\Delta v_o}{\Delta I_{ref}} = q [sI - A]^{-1} B. \quad (30)$$

3.4. VOLTAGE CONTROLLER DESIGN

A two-step controller can be implemented to address the issue of voltage regulation. The inner loop consists of a previously discussed current control loop, while the outer loop utilizes a PI controller, which is elaborated in this section. The analysis is performed using the small-signal method.

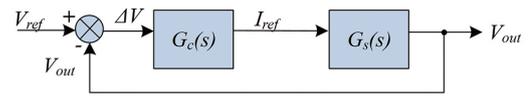


Fig. 4. Small signal closed-loop control diagram of the converter with PI controller

As depicted in Fig. 4, the PI controller is implemented between the GAM and frequency calculation block. $G_c(s)$ and $G_s(s)$ represents the transfer function of the controller and the converter, respectively. The measured output voltage V_{out} is subtracted from the desired output voltage V_{ref} . Then, the error signal is fed into a PI controller, which is implemented in closed-loop control.

The PI controller block determines the reference current as follows:

$$I_{ref} = K_p (V_{ref} - V_{out}) + K_i \int (V_{ref} - V_{out}) dt, \quad (31)$$

Where K_p and K_i are the proportional and integral coefficients of the PI controller, respectively.

The PI controller is used for voltage regulation. Combining the PI controller with an inner current loop reduces circulating reactive current during transients and provides alignment with secondary side voltage in most cases. The design and simulation results supporting these are presented in Section 5.

4. SIMULATION AND EXPERIMENTAL RESULTS

The converter with the proposed control has been implemented in MATLAB/Simulink. The main parameters used for converter simulation are given in Table 2. In the actual system, the input voltage is 400V, and the turn ratio n equals 3:1. However, for design simplicity, an input voltage of 30V and a turn ratio of 1:1 have been used.

Table 2. Simulation parameters

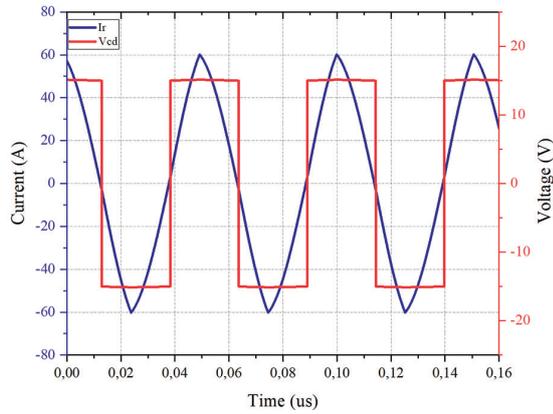
Parameter	Symbol	Value
Input voltage	V_{in}	35 V
Output voltage	V_{out}	20 V
Turn ratio	n	1:1
Resonant inductance	L_r	7.5 μ F
Resonant capacitor	C_r	15 μ F

4.1. EVALUATION OF SENSORLESS CURRENT CONTROL

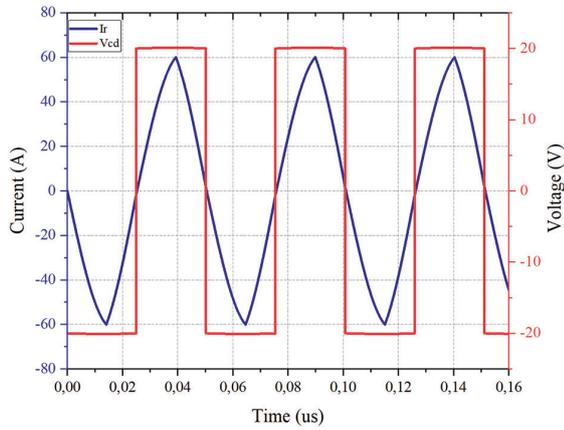
Evaluating performance and sensitivity under voltage, circuit parameters, and reference current variations helps analyze the proposed model's effectiveness.

To compare how changes in output voltage impact output aligned current, waveforms on secondary side voltage and resonant current for different output voltages can be analyzed. V_{out} is set to be equal (15; 20; 25) V at $I_{ref} = 50$ A. According to Fig. 5, it can be seen that the current is equal to zero at the beginning and the

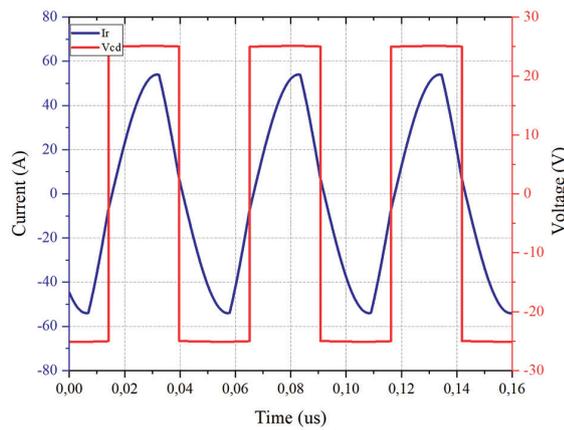
end of each switching cycle when $V_{out} = 15V$ and $20V$, ensuring ZVS and ZCS. When $V_{out} = 25V$, the current has shifted due to the frequency approaching the resonant frequency, resulting in inaccuracies.



(a)



(b)

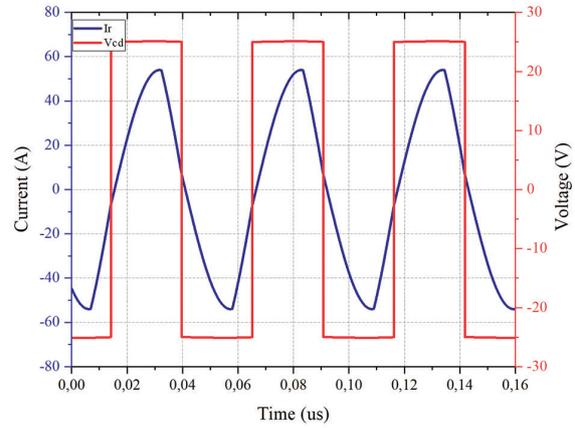


(c)

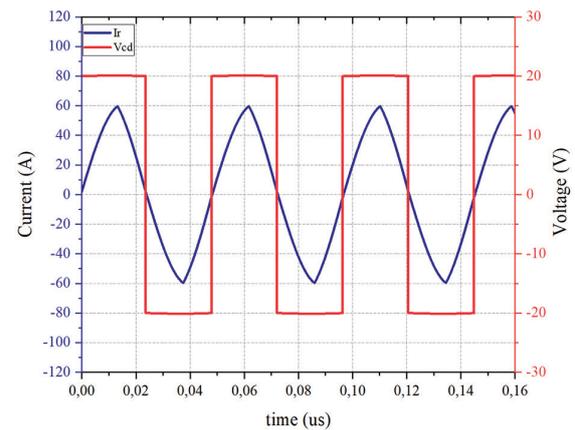
Fig. 5. Steady-state waveforms of the converter with proposed control technique @ $I_{ref} = 50A$ (a) $V_{out} = 15V$ (b) $V_{out} = 20V$ (c) $V_{out} = 25V$

To demonstrate the dependence of alignment on parameter variation, the steady-state waveforms of the circuit when the resonant capacitor is adjusted to be 10% higher and 10% lower than the actual capacitance C_r at constant values of $I_{ref} = 50A$ and $V_{out} = 20V$ are ob-

tained (Fig. 6). The results indicate that a 10% variation in capacitor value does not affect the ZVS and ZCS characteristics. Similarly, the steady-state waveforms when the resonant inductance is adjusted by 10% above and below the actual inductance L_r are also shown in Fig. 7. It demonstrates that change in inductance value results in a 2% variation in peak inductance current, with minimal effect on ZVS and ZCS characteristics.

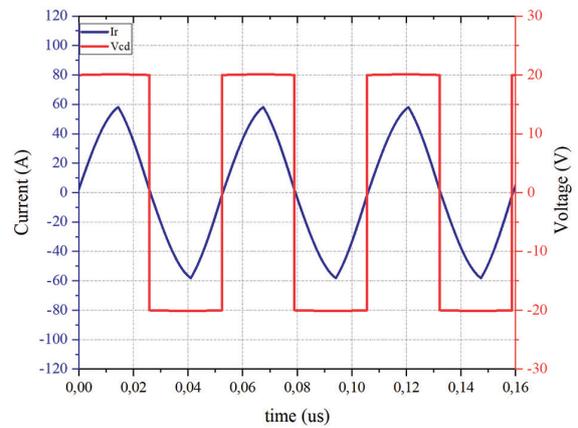


(a)

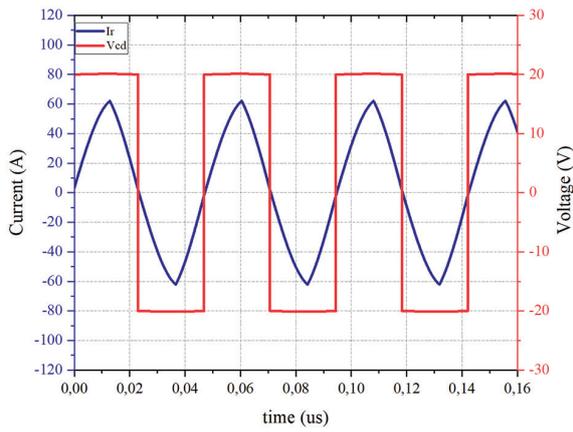


(b)

Fig. 6. Steady-state waveforms of the converter with proposed control technique @ $I_{ref} = 50A$, $V_{out} = 20V$ (a) $1.1 C_r$ (b) $0.9 C_r$



(a)



(b)

Fig. 7. Steady-state waveforms of the converter with proposed control technique @ $I_{ref} = 50A$ $V_{out} = 20V$ (a) $1.1 L_r$ (b) $0.9 L_r$

Reference current perturbation may affect output current alignment with output voltage. When the current is leading, the output current is greater than zero. In contrast, when the current is lagging, the output will drop below zero. The frequency response of the current control loop has a bandwidth of approximately 4 kHz. To understand the circuit's behavior near this point, the reference current was perturbed with a sine signal of 4 kHz and an amplitude of 3 A. The steady-state response of the output current under these conditions is shown in Fig. 8. The perturbations cause the current to fall below zero at the zero crossing, indicating that the inductance current is lagging. In the worst case, the power factor is 0.97 and in the best case, it is 0.99. However, this difference may not be visible in the simulation waveforms due to the small angle between the voltage and current. To better demonstrate the efficiency of the proposed system, switching and conduction losses are calculated.

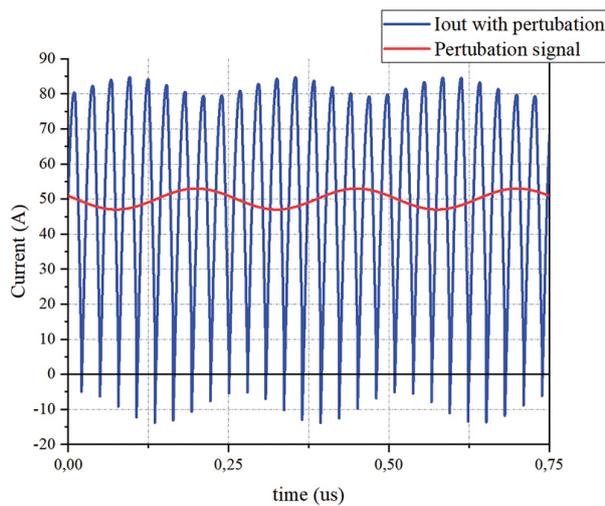


Fig. 8. Steady-state response of output current with reference current perturbation

$$P_{con} = I_{RMS}^2 R_{dc,ON} \quad (32)$$

ZVS switching occurs only on primary side lagging leg and its switching loss correspond to the internal capacitance C_{int} current flowing during switching $I_{turn-off}$ switching frequency f_s input voltage V_{in} and fall time t_{off} . Therefore, ZVS switching loss [33]:

$$P_{sw,ZVS} = C_{int} f_s I_{turn-off} V_{in} t_{off} \quad (33)$$

ZCS switching losses occur due to discharging of the parallel parasitic body capacitance of the transistors. It depends on internal capacitance C_{int} switching frequency f_s transistor voltages before turn-on which is equal to input voltage V_{in} for primary side losses and V_{out} for secondary side losses. Therefore, the loss during ZCS switching on primary side can be calculated as:

$$P_{sw,ZCS,pri} = \frac{1}{2} C_{int} V_{in}^2 f_s \quad (34)$$

Similarly, the loss during ZCS switching on secondary side:

$$P_{sw,ZCS,sec} = \frac{1}{2} C_{int} V_{out}^2 f_s \quad (35)$$

Total switching losses for circuit includes 2 ZVS and 2 ZCS switching on primary side, and 2 ZCS switching on secondary side. Therefore, total switching losses:

$$P_{sw} = 2P_{sw,ZVS} + 2P_{sw,ZCS,pri} + 2P_{sw,ZCS,sec} \quad (36)$$

These losses are calculated using parameters from the datasheets for both the primary side transistors INN100FQ025A [34] and secondary side transistors INN650TA030AH [35]. Fig. 9 illustrates the losses associated with the proposed strategy across various output currents while maintaining a steady output voltage of 20 V. The overall percentage of switching losses tends to be lower when the current is increased. Although higher current increases ZVS switching losses, it also leads to decreased switching frequency, which reduces both ZVS and ZCS switching losses. Additionally, conduction losses increase with current. Therefore, switching losses decrease with higher current. Overall, the losses remain below 8 W within the specified operational range, illustrating the efficient performance of the proposed modulation strategy across different operating conditions.

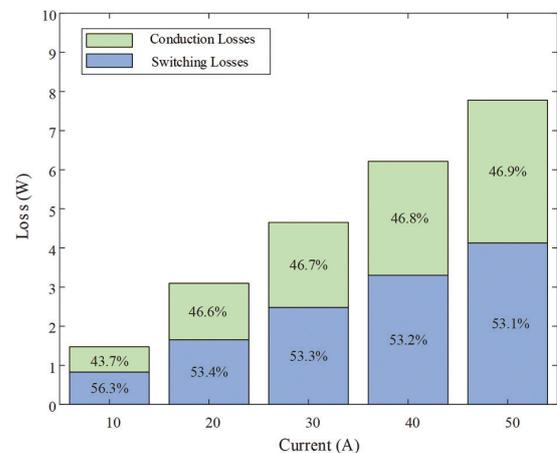


Fig. 9. Loss breakdown under output aligned current mode for different output current at output voltage of 20 V

4.2. EVALUATION OF CLOSED-LOOP VOLTAGE CONTROL

To evaluate outer loop voltage control, the frequency response characteristics of the current control loop with and without the outer loop can be analyzed. To achieve dynamic performance and stability, the gain and phase margins are set equal to 25 dB and 15°, respectively. Thus, using PI controller parameters from Table 2, the outer loop control satisfies targeted gain and phase margins, and it can be validated through the bode plot given in Fig. 10. Converter characteristics during a transient under load current step change of 10 A is shown in Fig. 11. Output current and voltage waveforms are smoothly adjusted to their new steady-state values.

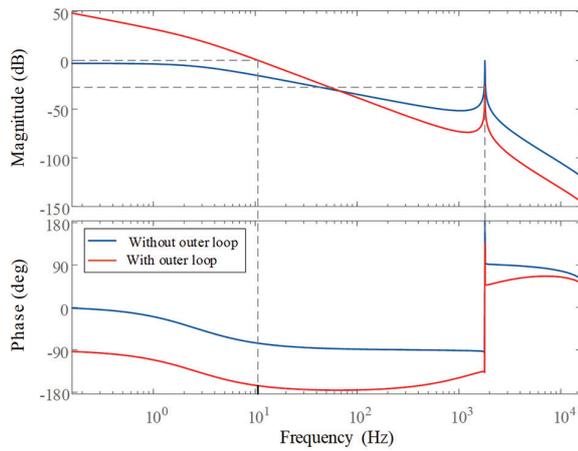


Fig. 10. Frequency response characteristics @ =50A without and with the outer loop

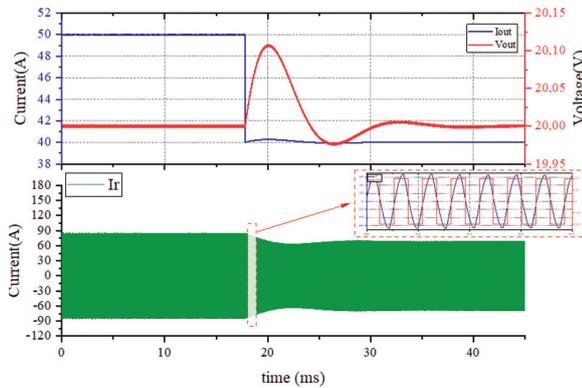


Fig. 11. Converter characteristics during transient under load current step change of 10A

Overall, analysis and simulation results represent that the proposed model provides current alignment with output voltage not only during steady-state but also during transients and parameter variation. The model performs satisfactorily even with frequency perturbations of up to approximately 4 kHz.

4.3. COMPARISON WITH EXISTING METHOD

A few papers have proposed a control technique for output-aligned current for resonant LC-DAB.

In this paper, the discrete-time technique has been implemented. This approach [8] is compared with the proposed method. This converter's parameters for open-loop modulation (phase shift ratio d and period T) can be determined using the reference output current I_{ref} along with the input voltage V_{in} and output voltage V_{out} , the resonant angle frequency ω_r . This method measures the resonant capacitor voltage over half of a sequence to facilitate accurate current calculations. The time t_2 is defined as half of the modulation period, while t_1 can be calculated using:

$$t_1 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad (37)$$

With coefficients:

$$a = \frac{-nV_{in}^2\omega_r^2}{C_r V_{out}(U_1 - U_2)} I_{ref},$$

$$b = -2U_1\omega_r^2, \quad (38)$$

$$c = \frac{nV_{in}\pi^2}{2C_r(V_{in} - V_{out})} I_{ref}.$$

The duty cycle can be derived using (16), and the switching frequency can be derived as follows:

$$f = \frac{d}{2t_1}, \quad (39)$$

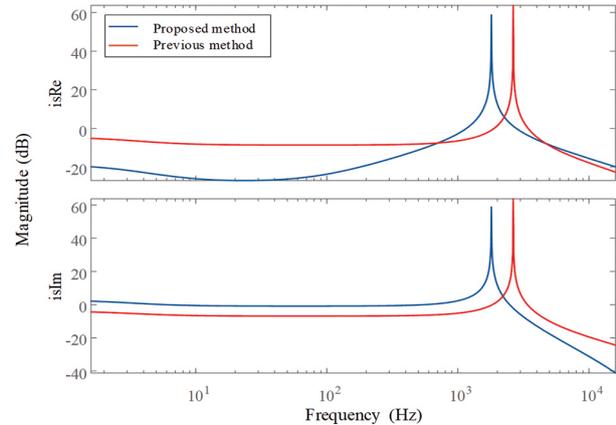


Fig. 12. Frequency response characteristics of real and imaginary parts of inductance current for two methods @ $I_{ref} = 50A$

Fig. 12 illustrates the frequency response of the real and imaginary parts of the inductor current using two methods. The existing model demonstrates a wider bandwidth range. However, the proposed method has better damping of the real part of the current over a wide range of frequencies, nearly up to 4 kHz. The proposed method has stable sensitivity close to the operating point, which minimizes the circulating output current even during transition. Fig. 13 shows the dependence of current on switching frequency between the previously established technique and the proposed method. The proposed model has lower switching frequencies for almost all current values, reducing switching and resistive losses.

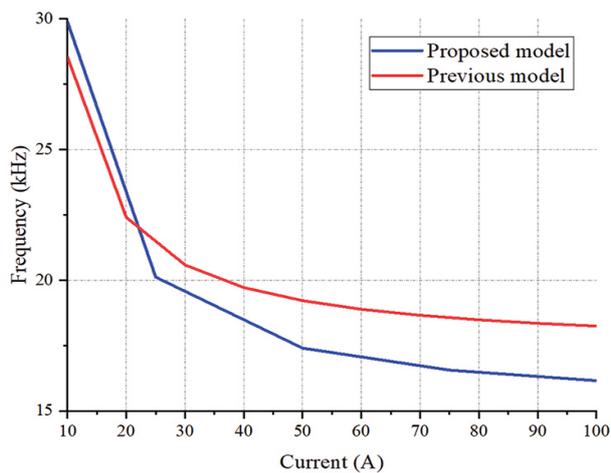


Fig. 13. Dependence of frequency on reference output current for two methods

In comparison with other works, the proposed technique has benefit of ability to work near resonant frequency which can further minimize switching losses. For instance, unified boundary control for resonant LC-DAB [26] can help to enhance converter efficiency by minimizing circulating current. However, it can work only at frequencies much higher than resonant frequency which may tend to higher switching losses. Another benefit of the proposed method is low cost and low complexity. In contrast, model with beat frequency dynamics [28] requires complex calculations to perform efficiently but has good performance in a wide frequency range. Moreover, the proposed model does not require the need of high frequency current measurement, as it is required by robust digital nonlinear controller [14]. While this controller can provide good dynamic performance with smooth transient response, it can only work with non-resonant DAB converter and requires high-frequency current measurement with complex calculations. As can be seen from extensive simulation, the proposed design also exhibits strong robustness, consistently maintaining current alignment with output voltage during steady-state conditions, transients, and parameter variations. In contrast, the accuracy of the method discussed in paper [21] relies on the model parameters, but it demonstrates good accuracy without the need for a PI-controller when the parameters are known.

Therefore, the proposed model is valuable for enhancing the dynamic performance of the resonant LC-DAB converter. It performs well under parameter variations, minimizes losses, effectively operates in the presence of a resonant circuit, and has low-cost with simple calculations.

4.4. EXPERIMENTAL RESULTS

The prototype of the LC-DAB was built and tested to demonstrate the effectiveness of the proposed model (fig. 14). The converter comprises four main components: the control PCB, the low-voltage side power PCB, the

high-voltage side power PCB, and a high-frequency transformer. The prototype was built on a printed board by using GANFET (INN100FQ025A) on the input side and GANFET (INN650TA030AH) on the output side. The control PCB utilizes a DSP (TMS320F28374D) alongside an FPGA (EPM570T100I). These three PCBs are interconnected through connectors in a sandwich configuration to facilitate PWM signal transmission. Both the DC and AC sides feature pads for transformer connections, positioned between the power and control boards to optimize space and enable easy replacement and inspection of all PCBs.

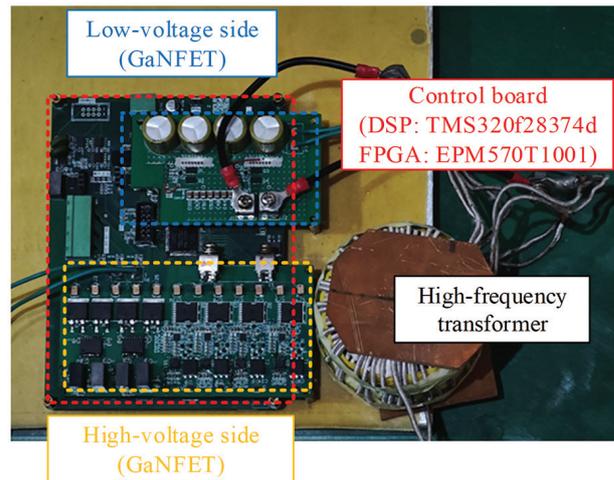
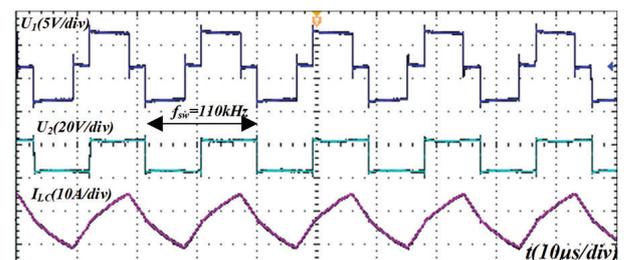
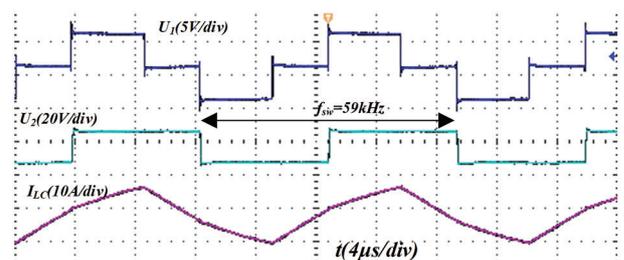


Fig. 14. Experimental setup

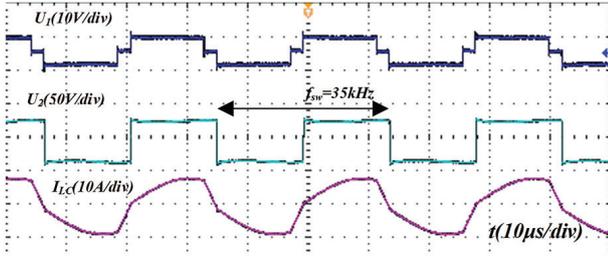
Fig. 15 depicts the experimental steady-state waveforms of primary and secondary side voltage and inductance current with proposed control technique at $I_{ref} = 10A$, $V_{in} = 35V$ for output voltages $V_{out} = 15V$, $V_{out} = 20V$, $V_{out} = 30V$. It represents alignment of output current with secondary side voltage even at frequencies near resonant. This shows the feasibility of proposed control technique.



(a)



(b)



(c)

Fig. 15. Steady-state waveforms of the converter with proposed control technique @ $I_{ref} = 10A$ $V_{in} = 35V$ (a) $V_{out} = 15V$ (b) $V_{out} = 20V$ (b) $V_{out} = 30V$

5. CONCLUSION

This work presents a model for the resonant LC-DAB with output-aligned current. The GAM method represents the model of the converter in a continuous form, making analysis more straightforward. This model employs a two-stage control loop: an inner loop for current control and an outer loop utilizing a PI controller for voltage control. This approach helps to minimize circulating current during switching at the output side of the converter. Moreover, the analysis and design process for the proposed control strategy is discussed, and the corresponding simulation studies are conducted in the MATLAB/Simulink environment along with experimental results are presented in order to validate the effectiveness of this scheme. The obtained results show that not only the proposed method provides a promising solution for improving the dynamic performance of DAB converters, but it also helps to analyze the converter's behavior in output-aligned current mode.

6. APPENDIX

A. 1

The matrices A, B, q of small-signal model, derived in (28)-(29) are listed in the following.

$$A = \begin{bmatrix} -\frac{R_s}{L_r} & k_3 + k_1' & 0 & -\frac{1}{L_r} & 0 \\ -k_3 - k_1 & -\frac{R_s}{L_r} & \frac{2}{\pi n L_r} & 0 & -\frac{1}{L_r} \\ 0 & -\frac{4}{C_{out} n \pi} & -\frac{1}{C_{out} R_{out}} & 0 & 0 \\ \frac{1}{C_r} & 0 & 0 & 0 & k_3 + k_1 \\ 0 & \frac{1}{C_r} & 0 & -k_3 - k_1 & 0 \end{bmatrix};$$

$$B = \begin{bmatrix} -i_{s_{Im}} k_1 - i_{s_{Im}} k_2 \\ i_{s_{Re}} k_1 + i_{s_{Re}} k_2 \\ 0 \\ -v_{s_{Im}} k_1 - v_{s_{Im}} k_2 \\ v_{s_{Re}} k_1 + v_{s_{Re}} k_2 \end{bmatrix};$$

$$q = [(0 \ 0 \ 1 \ 0 \ 0)];$$

With coefficients:

$$k_1 = \frac{2U_1 \sin\left(\frac{\pi U_2}{U_1}\right)}{\pi^2 I_{ref}^2 L_r};$$

$$k_2 = \frac{4U_1^2 \sin\left(\frac{\pi U_2}{U_1}\right)^2}{\pi^4 I_{ref}^3 L_r^2 \sqrt{\frac{1}{C_r L_r} + \frac{4U_1^2 \sin\left(\frac{\pi U_2}{U_1}\right)^2}{\pi^4 I_{ref}^2 L_r^2}}};$$

$$k_3 = \sqrt{\frac{1}{C_r L_r} + \frac{4U_1^2 \sin\left(\frac{\pi U_2}{U_1}\right)^2}{\pi^4 I_{ref}^2 L_r^2}}.$$

7. REFERENCES

- [1] F. Krismer, J. W. Kolar, "Accurate power loss model derivation of a high-current dual active bridge converter for an automotive application", IEEE Transactions on Industrial Electronics, Vol. 57, No. 3, 2010, pp. 881-891.
- [2] Y. Shi, R. Li, Y. Xue, H. Li, "Optimized operation of current-fed dual active bridge DC-DC converter for PV applications", IEEE Transactions on Industrial Electronics, Vol. 62, No. 11, 2015, pp. 6986-6995.
- [3] W. R. L. Garcia, P. Tixador, B. Raison, A. Bertinato, B. Luscan, C. Creusot, "Technical and economic analysis of the R-type SFCL for HVDC grids protection", IEEE Transactions on Applied Superconductivity, Vol. 27, No. 7, 2017, pp. 1-9.
- [4] A. Nisar, M. S. Thomas, "Comprehensive control for microgrid autonomous operation with demand response", IEEE Transactions on Smart Grid, Vol. 8, No. 5, 2017, pp. 2081-2089.
- [5] L. Zhuolan, Z. Yu, Z. Mengyuan, "Model and Control for Transient Process of Dual-Bridge Series Resonant Converter", IEEE Transactions on Power Electronics, Vol. 39, No. 8, 2024, pp. 9202-9220.
- [6] M. Sato, Y. Tada, M. Uno, "Multi-Port Converter Integrating Automatic Current Balancing Interleaved PWM Converter and Dual Active Bridge Converter with Improved Transformer Utilization", Proceedings of IECON 2020 - The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 18-21 October 2020, pp. 1377-1382.
- [7] X. Li, J. Huang, Y. Ma, X. Wang, J. Yang, X. Wu, "Unified modeling, analysis, and design of isolated bi-

- directional CLLC resonant DC-DC converters", *IEEE Journal of Emerging and Selected Topics in Power Electronics*, Vol. 10, No. 2, 2022, pp. 2305-2318.
- [8] G. Chen, N. Xu, L. Yuan, M. Humayun, M. M. Khan, "A DC-DC Center-Tapped Resonant Dual-Active Bridge with Two Modulation Techniques", *Electronics*, Vol. 9, No. 10, 2020, p. 1699.
- [9] S. M. S. I. Shakib, S. Mekhilef, "A Frequency Adaptive Phase Shift Modulation Control Based LLC Series Resonant Converter for Wide Input Voltage Applications", *IEEE Transactions on Power Electronics*, Vol. 32, No. 11, 2017, pp. 8360-8370.
- [10] P. Qu, Z. Li, P. Wang, Y. Li, "An improved analysis considering harmonic components for high-frequency isolated dual-active-bridge series resonant DC-DC converter", *Proceedings of the Transportation Electrification Asia-Pacific*, Beijing, China, 31 August - 3 September 2014, pp. 1-4.
- [11] E. Babaie, M. E. S. Mahmoodieh, M. Sabahi, "Investigating buck dc-dc converter operation in different operational modes and obtaining the minimum output voltage ripple considering filter size", *Journal of Power Electronics*, Vol. 11, No. 6, 2011, pp. 793-800.
- [12] E. Babaie, M. E. Seyed Mahmoodieh, H. M. Mahery, "Operational modes and output voltage ripple analysis and design considerations of buck-boost dc-dc converters", *IEEE Transactions on Industrial Electronics*, Vol. 59, No. 1, 2012, pp. 381-391.
- [13] S. P. Engel, N. Soltan, H. Stagger, R. W. De Doncker, "Dynamic and balanced control of three-phase high-power dual-active bridge DC-DC converters in DC-grid applications", *IEEE Transactions on Power Electronics*, Vol. 28, No. 4, 2013, pp. 1880-1889.
- [14] I. Askariyan, S. Bagawade, M. Pahlevani, A. M. Knight, A. Bakhshai, "Robust Digital Nonlinear Control System for Dual Active Bridge (DAB) DC/DC Converters With Asymmetric Half-Cycle Modulation", *IEEE Journal Of Emerging And Selected Topics In Industrial Electronic*, Vol. 1, No. 2, 2020, pp. 123-132.
- [15] S. Dutta, S. Bhattacharya, M. Chandorkar, "A novel predictive phase shift controller for bidirectional isolated DC to DC converter for high power applications", *Proceedings of the IEEE 30th International Symposium on Industrial Electronics*, Raleigh, NC, USA, 15-20 September 2012, pp. 418-423.
- [16] S. Dutta, S. Hazra, S. Bhattacharya, "A digital predictive current-mode controller for a single-phase high-frequency transformer-isolated dual-active bridge DC-to-DC converter", *IEEE Transactions on Industrial Electronics*, Vol. 63, No. 9, 2016, pp. 5943-5952.
- [17] S. Wei, Z. Zhao, K. Li, L. Yuan, W. Wen, "Deadbeat current controller for bidirectional dual-active-bridge converter using an enhanced SPS modulation method", *IEEE Transactions on Power Electronics*, Vol. 36, No. 2, 2021, pp. 1274-1279.
- [18] M. Nasr, S. Poshtkouhi, N. Radimov, C. Cojocaru, O. Trescases, "Fast average current mode control of dual-active-bridge DC-DC converter using cycle-by-cycle sensing and self-calibrated digital feed-forward", *Proceedings of the IEEE Applied Power Electronics Conference and Exposition*, Tampa, FL, USA, 26-30 March 2017, pp. 1129-1133.
- [19] N. Hou, Y. Zhang, Y. W. Li, "A Load-Current-Estimating Scheme with Delay Compensation for the Dual-Active-Bridge DC-DC Converter", *IEEE Transactions on Power Electronics*, Vol. 37, No. 3, 2022, pp. 2636-2647.
- [20] Y. Deng, W. Song, S. Yin, M. Zhong, K. Yang, X. Feng, "A Model Predictive Control Scheme without Current Sensor of Dual Active Bridge DC-DC Converters: Improving Dynamic Performance and Reducing Hardware Cost", *IEEE Transactions on Transportation Electrification*, Vol. 9, No. 2, 2022, pp. 2916-2928.
- [21] T. Q. Duong, S. J. Choi, "Sensor-Reduction Control for Dual Active Bridge Converter Under Dual-Phase-Shift Modulation", *IEEE Access*, Vol. 10, 2022, pp. 63020-63033.
- [22] L. Zhu, "A novel soft-commutating isolated boost full-bridge ZVS-PWM DC-DC converter for bidirectional high power applications", *Proceedings of the IEEE 35th Annual Power Electronics Specialists Conference*, Aachen, Germany, 20-25 June 2004, pp. 2141-2146.
- [23] R. Twiname, W. Malan, J. Minogue, D. J. Thrimawithana, U. K. Madawala, C. Baguley, "A novel

- dual active bridge topology with a tuned CLC network”, Proceedings of the IEEE International Conference on Industrial Technology, Busan, South Korea, 26 February - 1 March 2014, pp. 895-900.
- [24] L. Rossetto, G. Spiazzi, “Series resonant converter with wide load range”, Proceedings of the IEEE Industry Applications Conference Thirty-Third IAS Annual Meeting, St. Louis, MO, USA, 12-15 October 1998, pp. 1326-1331.
- [25] D. Czarkowski, M. K. Kazimierczuk, “Phase-controlled series-parallel resonant converter”, IEEE Transactions on Power Electronics, Vol. 8, No. 3, 1993, pp. 309-319.
- [26] G. Chen, X. Li, S. Zhou, “Unified Boundary Control with Phase Shift Compensation for Dual Bridge Series Resonant DC-DC Converter”, IEEE Access, Vol. 8, 2020, pp. 131137-131149.
- [27] D. J. Linton, C. A. Teixeira, R. H. Wilkinson, B. P. McGrath, D. G. Holmes, J. Riedel, “Adaptive modulation of resonant DAB converters for wide range ZVS operation with minimum reactive circulating power”, IEEE Transactions on Industry Applications, Vol. 58, No. 6, 2022, pp. 7396-7407.
- [28] S. Chakraborty, K. Hatua, “Modeling with beat frequency dynamics and phase-frequency control design for a dual-bridge series resonant converter”, IEEE Transactions on Power Electronics, Vol. 69, No. 8, 2022, pp. 7952-7962.
- [29] T. Chen, R. Yu, A. Q. Huang, “A bidirectional isolated dual-phase-shift variable-frequency series resonant dual-active-bridge GaN AC-DC converter”, IEEE Transactions on Industrial Electronics, Vol. 70, No. 4, 2023, pp. 3315-3325.
- [30] H. Qin, J. W. Kimball, “Generalized Average Modeling of Dual Active Bridge DC-DC Converter”, IEEE Transactions on Power Electronics, Vol. 27, No. 4, 2012, pp. 2078-2084.
- [31] A. K. Dubey, N. Lakshminarasamma, “Modeling of series resonant dual active bridge converter for dc microgrid application”, Proceedings of the IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems, Prague, Czech Republic, 28 June - 1 July 2022, pp. 1-6.
- [32] K. Takagi, H. Fujita, “Dynamic control and performance of a dual-active-bridge DC-DC converter”, IEEE Transactions on Power Electronics, Vol. 33, No. 9, 2017, pp. 7858-7866.
- [33] R. Zhang, M.M. Khan, J. Zhao, A. Mukhametdinova, M. Q. Khan, “A Four-Degrees-of-Freedom Modulation Strategy for Series Resonant DAB DC-AC Application”, International Journal of Circuit Theory and Applications, 2024. (in press)
- [34] Innoscience, “100V GaN Enhancement-mode Power Transistor”, INN100FQ025A datasheet, https://www.innoscience.com/site/product_details/479 (accessed: 2025)
- [35] Innoscience, “650V GaN Enhancement-mode Power Transistor”, INN650TA030AH datasheet, https://www.innoscience.com/site/product_details/830 (accessed: 2025)

Stream-based Identification of Gender using Noninvasive Electroencephalographic Technology

Original Scientific Paper

Bat-Erdene Gotov

Mongolian National University of Medical Sciences,
Department of Basic Science
Ulaanbaatar, Mongolia
baterdene.g@mnums.edu.mn

Tengis Tserendondog*

Mongolian University of Science and Technology,
Department of Electronics
Ulaanbaatar, Mongolia
tengis@must.edu.mn

Uurtsaikh Luvsansambuu

Mongolian University of Science and Technology,
Department of Electronics
Ulaanbaatar, Mongolia
uurtsaikh@must.edu.mn

*Corresponding author

Munkhbayar Bat-Erdene

Mongolian University of Science and Technology,
Department of Information Networks and Security
Ulaanbaatar, Mongolia
munkhbayar.b@must.edu.mn

Batmunkh Amar

Mongolian University of Science and Technology,
Department of Electronics
Ulaanbaatar, Mongolia
abatmunkh@must.edu.mn

Dongsung Bae

Sangmyung University,
Department of Software
Seoul, Republic of Korea
paeds915@smu.ac.kr

Abstract – Numerous studies on EEG signals have revealed differences in brain activity patterns between males and females. However, these differences aren't always consistent or significant, as they can be affected by factors like age, task engagement, and specifics of EEG measurements. In our research, we introduce a new approach to detect gender called 'Stream-based Identification of Gender using Noninvasive Electroencephalographic Technology'. We employed this technique to investigate how male and female brains respond differently during video streaming tasks with the aim of exploring functional disparities between them. This study aims to advance our understanding of gender-specific brain responses. We used data collected in our previous research from 122 volunteers (85 male, 37 female). Utilizing a deep learning (DL) approach allowed us to achieve 99% accuracy in gender identification. The applications of our model extend to various fields, including advertisements, multi-level security systems, and healthcare, showcasing the potential of advanced machine learning techniques in neuroscientific research.

Keywords: EEG, sex, gender difference, machine learning, deep learning

Received: October 4, 2024; Received in revised form: December 9, 2024; Accepted: January 10, 2025

1. INTRODUCTION

Exploring gender differences can yield crucial insights across diverse fields, including healthcare, cognitive science, education, and sociology. Here are several reasons why identifying and understanding gender differences is important.

Health and Medicine - Men and women respond differently to medications due to their unique biology [1]. Understanding these differences helps doc-

tors provide better treatments, improving results and reducing side effects. Diseases like heart disease may show different symptoms in men and women, so research helps doctors diagnose and treat them more effectively for each sex.

Technology and AI - Biometric technologies like facial recognition and voice ID work by analyzing physical traits that vary between sexes [2]. Understanding these differences makes the systems more secure and user-friendly. Designing technology that considers

gender differences can improve user satisfaction by accounting for physical and behavioral preferences.

Education and Workplace - Males and females often have different learning styles [3]. Adapting teaching to these differences can improve learning. In workplaces, men and women may face unequal opportunities for career growth and pay. Research helps address these issues by promoting policies like diversity training and flexible work option.

Social Understanding and Policy - Studying gender differences helps challenge stereotypes and shows how gender influences behavior and roles in society [4]. This leads to greater respect and acceptance of diverse gender identities.

Understanding these differences helps create more inclusive and effective practices that benefit all individuals, regardless of gender. There are three main approaches to identifying gender or sex differences:

Biological Methods: Genetic Testing, analyzing DNA to determine sex chromosomes or genetic variations related to biological sex. Hormonal Analysis, measuring hormone levels to understand physiological differences between sexes. Physical Examination, assessing anatomical features and secondary sexual characteristics to determine biological sex.

Behavioral and Psychological Methods: Self-Report, individuals self-identifying their gender through surveys, interviews, or questionnaires based on their internal sense of identity. Psychological Assessment, evaluating gender identity, expression, and related behaviors through psychological tests or clinical evaluations.

Technological and Computational Methods: Biometric Analysis, using biometric data (e.g., facial recognition, voice patterns) to infer gender based on physical characteristics. Machine Learning Algorithms, training algorithms on datasets containing biometric or behavioral data to predict or classify gender.

Biometric data that can identify sex or gender differences typically include physiological, anatomical, and sometimes behavioral characteristics that differ between males and females. Here are some common types of biometric data used for sex or gender identification. Differences in facial structure, such as jawline, cheekbones, and distance between eyes, can be analyzed using facial recognition technology [5, 6]. Acoustic properties of speech, including pitch, frequency, and resonance, are used in voice analysis to distinguish between male and female voices [7, 8]. While fingerprints themselves are not inherently gender-specific, certain patterns or ridge densities may correlate with biological sex [9]. Measurements of hand size, finger length ratios, and palm characteristics can be analyzed to infer sex differences [10]. Unique patterns of blood vessels in the retina can be scanned and compared for differences between

males and females [11]. Like retinal patterns, unique iris patterns can be scanned and compared for sex-related differences [12]. Genetic markers, including those found on sex chromosomes can definitively determine biological sex [13]. Differences in walking patterns and movements can sometimes indicate gender-specific characteristics, though this is less commonly used compared to other biometric methods [14]. The shape and size of the ear, including the earlobe and inner ear structures, can be analyzed for gender identification [15].

Brain Structure: While not typically used in everyday biometric applications, differences in brain structure and function have been studied to understand gender differences in cognition and behavior. Research has identified several differences in brain structure between genders, although these differences exist on a spectrum and can vary widely among individuals. Here are some key findings.

On average, male brains are larger than female brains, with 8% to 13% more total volume [16]. Males generally have bigger cerebrum, cerebellum, amygdala, and hippocampus. However, brain size doesn't determine intelligence or cognitive abilities.

Females tend to have more grey matter compared to white matter than males [17]. Grey matter processes information, while white matter connects different brain regions.

There may be slight differences in the thickness and surface area of the cerebral cortex between males and females, which could affect cognitive abilities and emotional processing [18].

The hippocampus, key for memory, and the amygdala, linked to emotions, can differ in size and connectivity between males and females, possibly impacting learning and emotional regulation [19].

The corpus callosum, which connects the brain's two halves, tends to be larger in females [20]. Both genders show neuroplasticity, but this adaptability is influenced by hormones and life experiences.

Functional imaging shows differences in brain activity during tasks between males and females, reflecting different cognitive strategies. While there are average differences in brain structure, individual variation is significant. Genetics, hormones, and experiences all contribute to these differences. Ongoing research aims to better understand these complexities.

In our study, we use Electroencephalograph (EEG) signals to identify gender differences, but there are many studies based on EEG data for various types of classification, such as emotion recognition [21], cognitive load [22], sleep stages [23], mental disorders [24], motor imagery [25], attention level, fatigue detection [26], seizure detection [27], speech imagery [28], drug effect [29], learning disabilities [30], pain detection [31], and more. Determining gender from

EEG signals alone is not straightforward and generally not highly accurate. EEG records the electrical activity of the brain and reflects complex neural patterns associated with cognitive and physiological activities. While there are broad generalizations about gender differences in brain structure and function, EEG signals are typically used to study brain states, neural disorders, or cognitive function rather than to identify biological characteristics like gender.

Research into gender differences in EEG has shown some variance between males and females in aspects such as amplitude and frequency of the brain waves. For instance, studies have suggested differences in the alpha wave frequency (a type of brain wave commonly associated with relaxation and lack of visual stim-

uli), with women generally showing higher frequencies than men. However, these differences are often subtle and influenced by a range of factors including age, hormonal status, and health conditions. In Table 1, a research review is shown on gender classification based on EEG signals using various methods.

The first section of this paper introduces the importance of gender classification and the different types of classification. Section 2 covers the data collection process, preprocessing, and presents some classification techniques based on EEG characteristics. Section 3 details our proposed deep learning method for gender classification using video streaming data. Finally, Section 4 provides the conclusion.

Table 1. Methods and Findings of similar studies

Ref.	Device	Number of Electrodes and electrode placement	Participants	Age	Hand preference, eyes open, eyes close	Stimuli	Preprocess
32	EEG-4418, Nihon Kohden	18, (10-20)	40 (20 males)	19-26	Right-handed, EC	Rest and During Photic Stimulation	Bandpass filter 0.3-60 Hz
33	Brain-tronics	30, (15% extending, distance to 5% anterior of Fz)	20 (10 males)	mean 24.5	Right-handed, EO	Visual based anagram and mental arithmetic tasks	Bandpass filter 1-30 Hz
34	Nihon Kohden	19, (10-20)	30 (15 males)	20-30	Right-handed, EO	Mental rotation task Shepard-figures.	Bandpass filter
35	EEG-16 S, Medikor, Hungary	16, (10-20)	30 (15 males)	19-23	Right-handed, EC	Listening test tapes and memorize words	Bandpass filter between 0.3-30 Hz
36	Quick-Cap	19, (10-20)	76 (38 males)	mean 21	Right-handed, EO	Emotional intelligence test WAIS-R and MSCEIT	Bandpass filter 0.15-50.0 Hz
37	Electrical Geodesics	128, (geodesic sensor net)	114 (54 males)	18-30	Right-handed, EO	Eriksen Flanker Task with arrow stimuli presented	Low-pass filtered at 30 Hz
38	BrainAmp Standard	37, (10-20)	27 (13 males)	mean 24.6	Right-handed	Spatial navigation in virtual environments	0.5 -50 Hz filter.
39	PSYLAB	6, (Pseudo-unipolar recordings were acquired)	42 (21 males)	20-29	EO, EC	Visual and motion-onset stimulation	Frequency band of 0.3-100
40	Australian EEG Database	23, (10-20)	40 (20 males)	19-69	EO, EC	Resting state	1 -30 Hz filter.
41	Emotiv Epoc	14, (10-20)	60 (35 males)	6-55	EC	Resting position	DWT
42	Neuroscan	30, (10-20)	28 (13 males)	18-30	EO, EC	Resting conditions	Bandpass filter 0.15-45 Hz
43	Emotiv Epoc	14, (10-20)	60 (35 males)	6-55	EC	Resting position	DWT
44	Biosemi ActiveTwo system	32, (10-20)	32 (16 males)	19-37	EO	Watching the audiovisual clips,	Bandpass filter 4-45 Hz
45	EEGLab MATLAB	19, (10-20)	134 (41 males)	mean 46	EC	The patients received 4 weeks of antidepressant treatment. Acoustic stimuli	PREP pipeline
46	Geodesics system	64, (Geodesics system)	61 (30 males)	mean 12.48	EC	Sleep condition	Bandpass filtered signal from the Hilbert transform
47	Emotiv Epoc	14, (10-20)	10 (6 males)	22.6 ± 2.75	EO	Short video clips with audio	Bandpass filter 0.5-64 Hz
48	Neuroscan Synamps2	30, (10-20)	80 (40 males)	18-26	Right-handed, EO, EC	Resting conditions	DFT
49	USBamp, Austria	16, (10-20)	20 (10 males)	22-30	Right-handed, EO	Color and black/white stimuli	Bandpass filter 0.1-50 Hz
50	Nicolet One EEG System	21, (10-20)	1140 (504 males)	18-88	EC, EO	From clinical recordings and exhibited a mix of paradigms with resting states, stimuli, and so forth.	Bandpass filter 1-40 Hz
51	BioSemi ActiveTwo	64, (10-20)	227 (145 males)	20-77	EC, EO	Resting state	Bandpass filter 2-20 Hz

2. DATA COLLECTION AND PREPROCESSING

In our study, we included students and teachers from the Mongolian University of Science and Technology as participants, with a total of 122 individuals comprising 85 men and 37 women. Among them, 112 participants were aged between 17 and 23 years, while the remaining 10 were aged between 25 and 54 years. This age distribution was considered when analyzing gender differences in EEG signals, as age related variations in brain activity could potentially influence the results. We controlled for age related variability by ensuring that the majority of participants were within a relatively narrow age range (17–23 years), reducing the impact of age on the observed gender differences. This approach allowed us to focus on the primary objective of investigating gender specific EEG patterns while accounting for potential confounding effects of age. Furthermore, we ensured that age groups were balanced across genders to further minimize any potential bias in the results.

We acknowledge that the gender imbalance in our sample may affect the generalizability of our findings and that a more balanced gender representation could provide additional insights. The gender distribution in our study was primarily influenced by the availability and recruitment process of participants. Despite efforts to recruit a diverse sample, logistical constraints, such as the availability of volunteers and their willingness to participate, resulted in an unequal representation.

Although there was a gender imbalance, we implemented measures to minimize bias in the analysis. For

instance, data preprocessing, feature extraction, and modeling were conducted without gender based disparities. Separate analyses were performed for male and female participants, revealing consistent trends in EEG features across both groups. These trends suggest that our findings reflect generalizable patterns. However, we recognize that the smaller female sample size may limit the detection of certain subtle gender specific differences.

To address this limitation, we plan to conduct follow up studies with a more balanced gender representation to validate and extend our findings. This will enable us to assess whether the observed patterns remain consistent across a more representative population.

Before starting brain signal measurements, they filled out a questionnaire, and their psychological status was rated on a scale of 1 to 10. Psychological status will be used in further research to determine the individual's condition. Participants are allowed to move their heads, facial muscles, eyes, blink, mouth, hands, and make slight body movements. In this measurement, the participant sits in front of a screen and watches a 120-second silent video featuring nature, cars, animals, and male and female actors, as shown in Figure 1 below. The video has no sound and features a very colorful nature, with land and sea animals, and includes very famous male and female actors. In some scenes, such as those involving cars and motorcycles, rapid changes occur, including with female actors. The nature and animal scenes are relaxing.

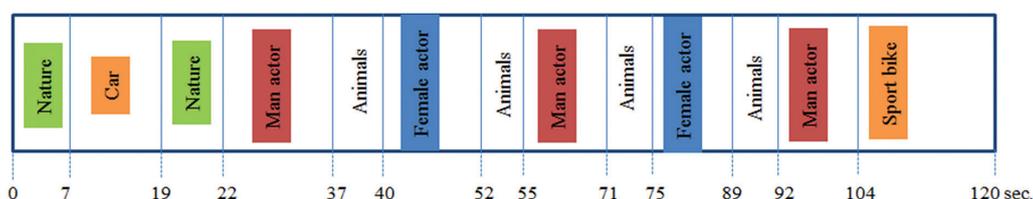


Fig. 1. Processes in the video

Testing was conducted using the Emotiv EpocX 14-channel headset, where data was recorded at a resolution of 16 bits and a sampling rate of 128 samples per second. In this study, electrodes (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, plus two additional CMS/DRL and P3/P4 electrodes) were placed according to the 10-20 international standard on the scalp. All channels were retained for analysis without exclusion. Due to the unequal number of male and female participants, we balanced the groups with 37 women and 37 randomly selected men. The training set consisted of 31 women and 31 men, while the test set included the remaining 6 women and 19 men.

In all subsequent experiments, a bandpass filter ranging from 0.1 Hz to 40 Hz was applied to remove inherent noise from the EEG signal recordings, ensuring improved signal clarity. Inconsistencies in gender differences in EEG signals were addressed through

rigorous preprocessing, task standardization, and balanced sampling. Participants performed a standardized video streaming task under controlled conditions to minimize external influences such as noise and light. Data collection was conducted at similar times of day to account for circadian rhythms, and participants were screened for neurological conditions, fatigue, and medication use.

To ensure consistent task engagement, participants were provided with clear instructions and regular breaks to prevent fatigue during the EEG recording sessions. Additionally, task engagement was monitored through behavioral observations and self-reports collected after each session. While task engagement was not directly quantified, the experimental design was structured to maintain high and consistent levels of attention among participants.

Epochs containing artifacts were excluded from the analysis, and only high-quality EEG recordings with minimal artifacts were included. Equal sampling was ensured during training to mitigate gender imbalances in the dataset.

We recorded brain signals from a 120-second video clip, and the total recording length was trimmed to 117 seconds to standardize the data. Therefore, the size of one person's brain recording is 14 x 14976 data points (channels by 117 seconds * sampling rate). Afterwards, we computed the mean of brain recordings from all

male participants and separately computed the mean of brain recordings from all female participants. This means that the resulting evoked object contains the averaged signal for each of the 14 channels over the specified time window. Based on previous studies, we employed Independent Component Analysis (ICA) to remove eye-related noise from each participant's brain signal recordings. This method allows for the separation of eye movement artifacts from the desired brain signals, improving the quality of data analysis and interpretation.

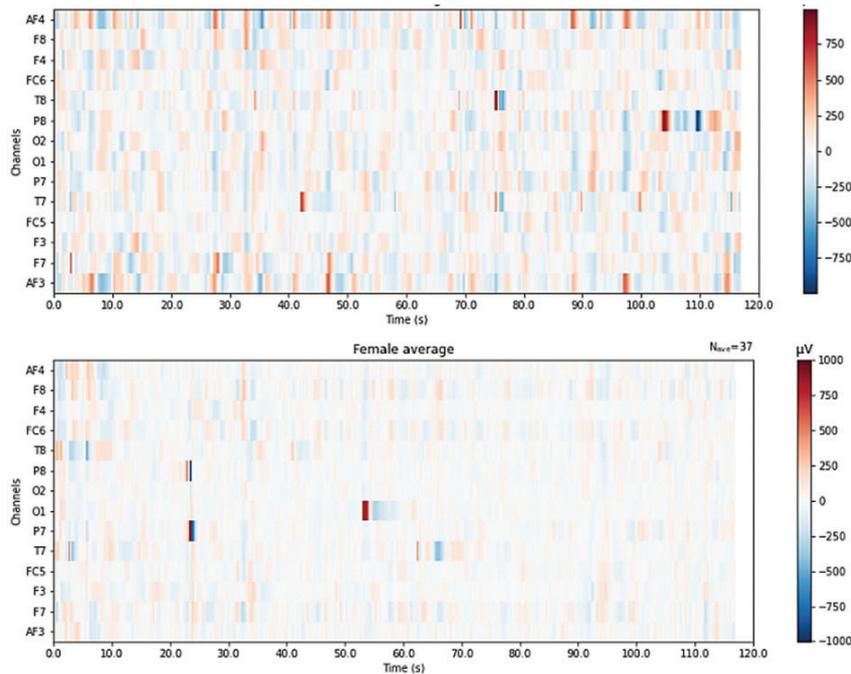


Fig. 2. Average EEG recordings of men's and women's brains

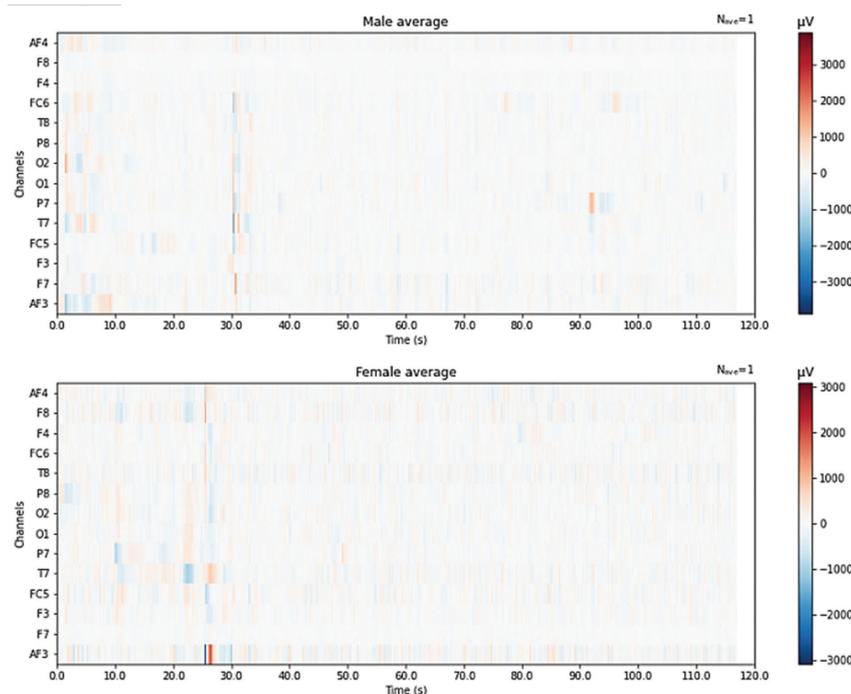


Fig. 3. Average EEG recordings of men's and women's after applying ICA

As the last step in data preparation, we performed gender-specific signal normalization. This involved subtracting the average EEG signal of females (used to create a 'common pattern' for female-EEG) from each male participant's EEG data and subtracting the average EEG signal of males (used to create a 'common pattern' for male-EEG) from each female participant's EEG data. This procedure ensured that gender-specific differences in baseline EEG activity were minimized before further analysis.

When we subtract the average female EEG data from the first male participant's EEG data, the result highlights the differences between this specific male's EEG signals and the typical female EEG signals. For each channel and each time point, this difference signal shows how the first male's EEG data deviates from the average female EEG data. Positive values indicate that the first male's EEG activity at those points is higher than the average female EEG activity. Negative values indicate that the first male's EEG activity at those points is lower than the average female EEG activity. This result provides a relative comparison between the first male's EEG and the average female EEG. It can be used to study gender-related differences in brain activity or to identify specific features that distinguish the first male's EEG pattern from the

average pattern observed in females. By examining the difference signal, we can identify unique patterns or anomalies in the first male's EEG data compared to the average female EEG data. Our dataset comprises EEG data from male participants (37 subjects, 14 channels, 14,976 time points) and female participants (37 subjects, 14 channels, and 14,976 time points).

Before proceeding with classification, we performed wavelet transformation on the average EEG data separately for male and female participants. This transformation was conducted to analyze EEG signals in both time and frequency domains, allowing us to detect and extract specific temporal and spectral features that might distinguish between genders. By applying wavelet transformation, we aimed to uncover nuanced variations in brain activity patterns across different frequency bands and time intervals, thereby enhancing our ability to discern gender-specific neural signatures in the EEG recordings.

AF3 and AF4 electrodes are located on the left and right sides of the forehead, near the front, and are typically used to monitor activity in the prefrontal cortex, which is associated with higher cognitive functions and emotional regulation. EEG studies involving AF3 often focus on attention, executive function, and emotional processing.

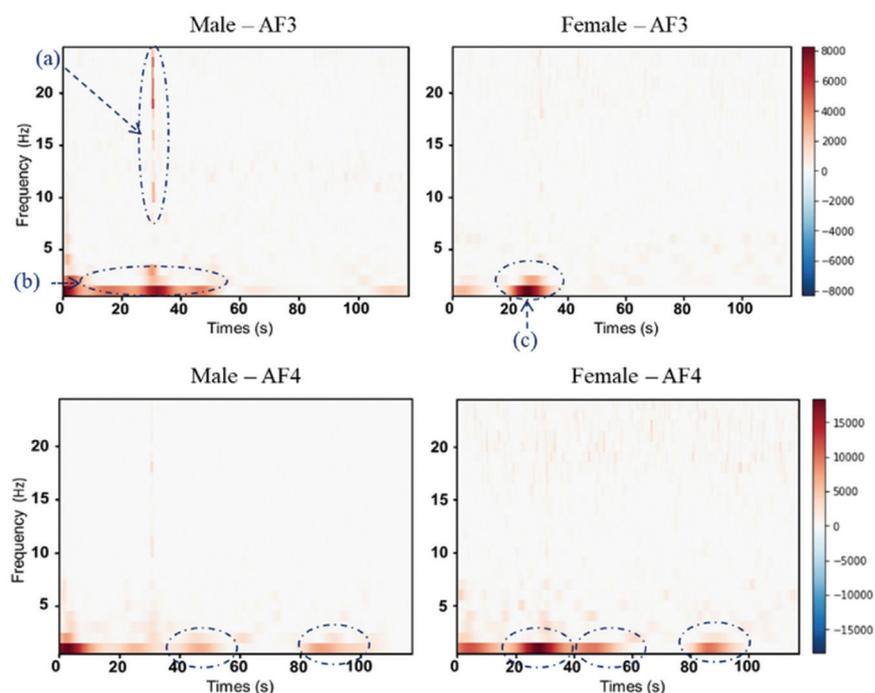


Fig. 4. AF3 and AF4 electrodes in males and females

(a) - between 27 and 30 seconds, high frequencies become active at the AF3 electrode on the man's head. This coincides with the "text" appearing in the video we showed, which the participants may have focused on. (b) - car, nature and man and female actor was shown. (c) - in the video between 22 and 37 seconds, a male actor was shown, and there was activity observed at the

female AF3 point. In AF4, there was activity in low frequencies observed when both male and female actors were shown during the video stream.

F7 and F8 are located on the left and right sides of the head, in front of the ears, and are used to study brain activity related to language processing, emotional responses, and other cognitive functions.

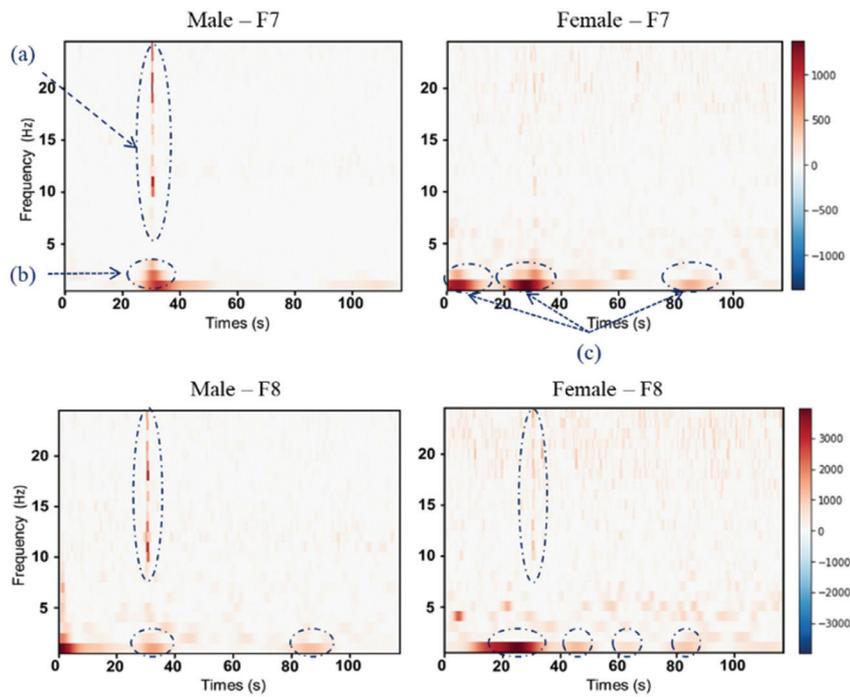


Fig. 5. F7 and F8 electrodes in males and females

(a) – The same high-frequency activity was observed in F7 when the “text” appeared in the video. (b) – a male actor was shown. (c) - male and female actors were shown in the video. In male F8, low-frequency activity was observed when the female actor was shown in the video. In female F8, low-frequency activity was

observed when the male actor was shown in the video.

F3 and F4 are located on the left side of the head, in the frontal lobe, and they are used to study brain activity related to cognitive functions such as attention, problem-solving, and motor control, as well as emotional processing.

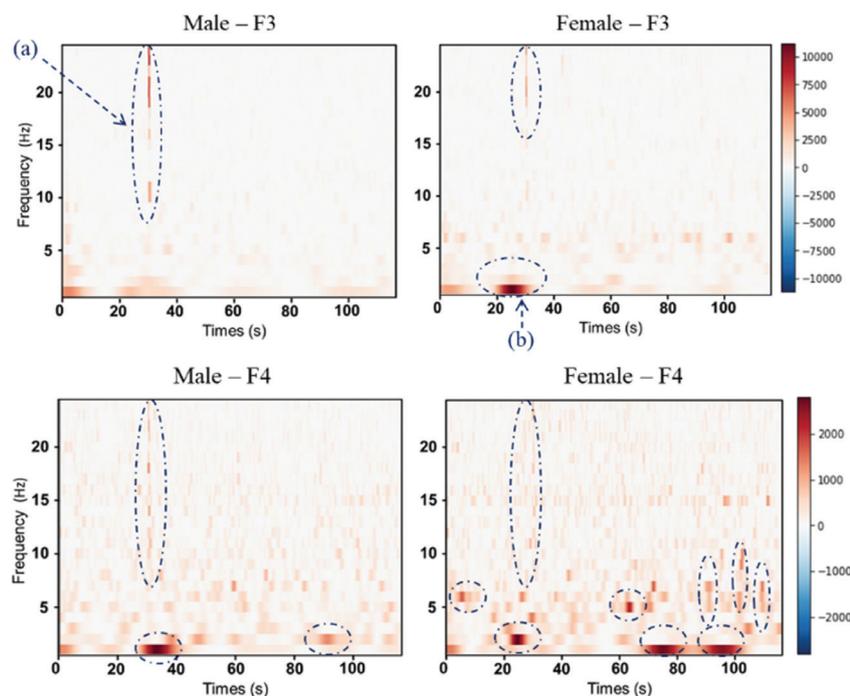


Fig. 5. F3 and F4 electrodes in males and females

In both male and female F3 and F4, high frequencies were still observed when text appeared in the video. In low frequencies, there were activities shown in actors

of the opposite gender. In female F4, activities were observed in the frequency range between 4-12 Hz when the opposite gender appeared in the video.

FC5 and FC6 are located on the left side of the head, in the frontal-central region, and are used to study brain

activity related to cognitive processes such as working memory, decision-making, and language production.

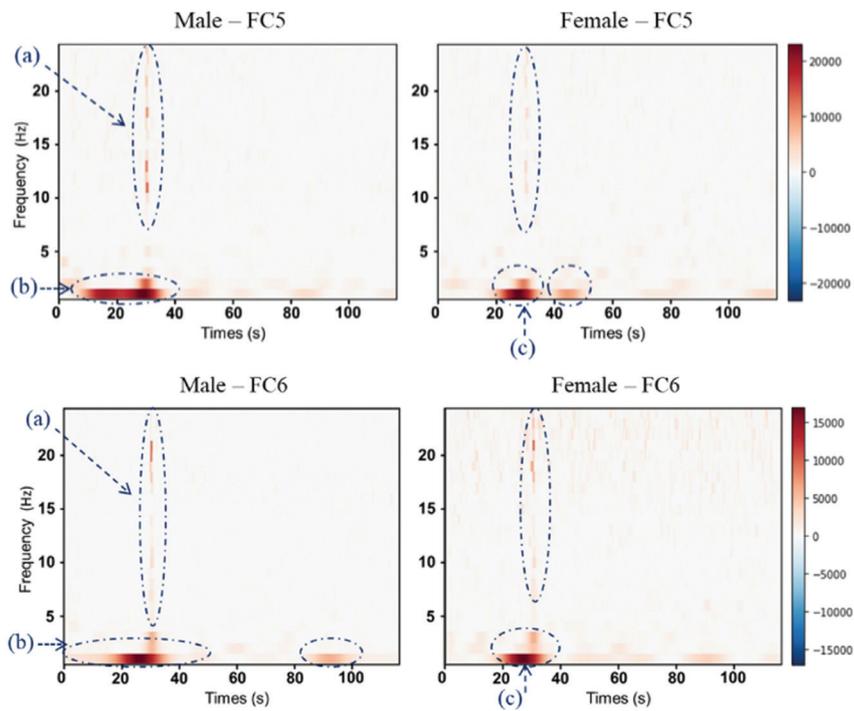


Fig. 7. FC5 and FC6 electrodes in males and females

In FC5 and FC6, elevated high-frequency activity was detected when text appeared in the video, whereas increased low-frequency activity was noted when actors were shown in the video. T7 and T8 are actually located

on the left and right sides of the head, respectively, in the temporal lobe. They are used to study brain activity related to auditory processing, language comprehension, and memory functions.

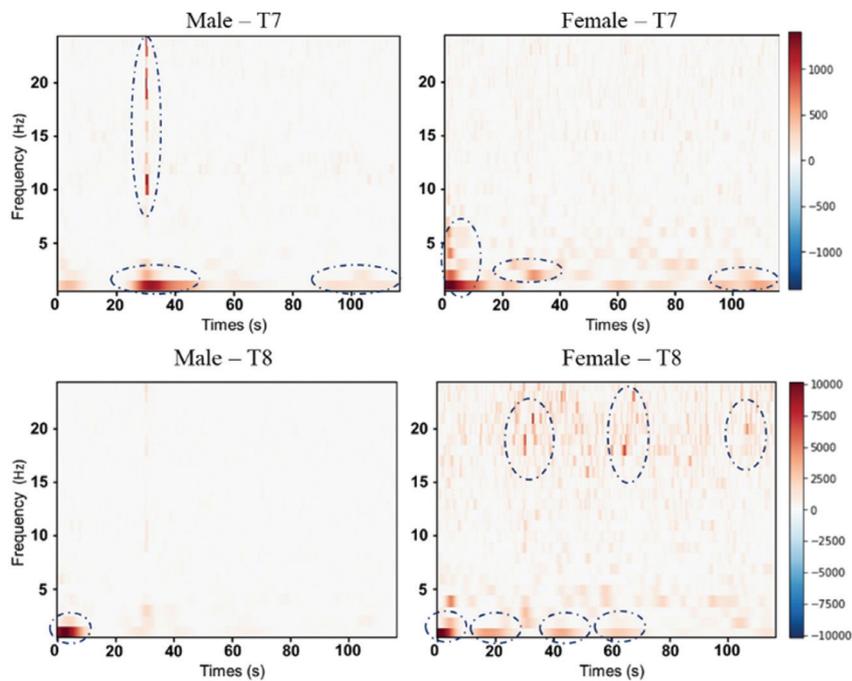


Fig. 8. T7 and T8 electrodes in males and females

In male T7, high-frequency activity can be observed when text appears in the video, whereas low-frequency activity is present at the start of the video. In female T8, there are also instances of high-frequency activity.

P7 and P8 are located on the left side of the head, in the parietal lobe, and are used to study brain activity related to sensory processing, spatial awareness, and attentional processes.

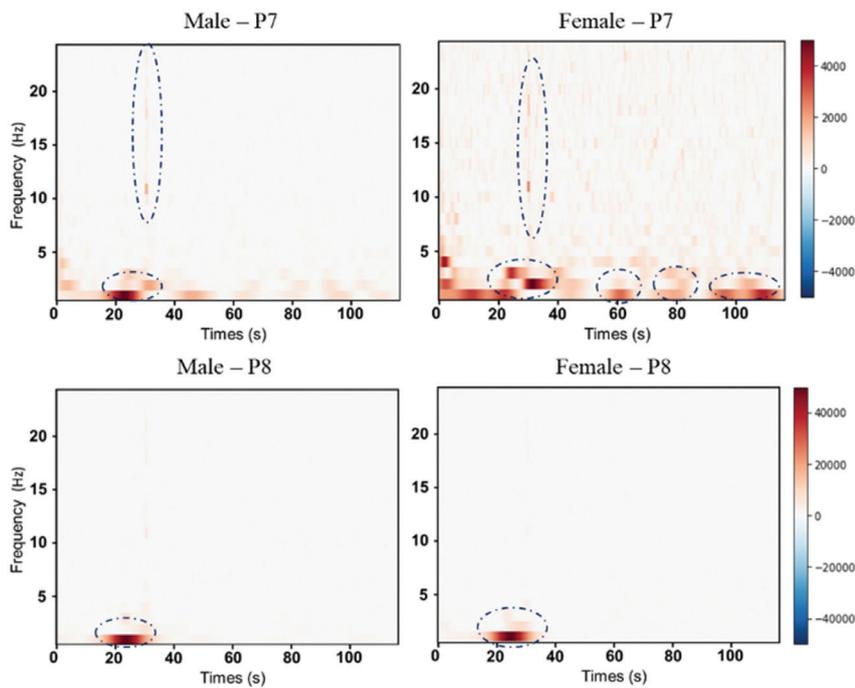


Fig. 9. P7 and P8 electrodes in males and females

In female P7, high-frequency activities are observed along with some lower frequency activities when actors are shown in the video. In both male and female P8, only one instance of low-frequency activity is observed during the same time period.

O1 and O2 are actually located on the left and right sides of the head, respectively, in the occipital lobe. They are indeed used to study brain activity related to visual processing, including visual perception, attention to visual stimuli, and various aspects of visual cognition.

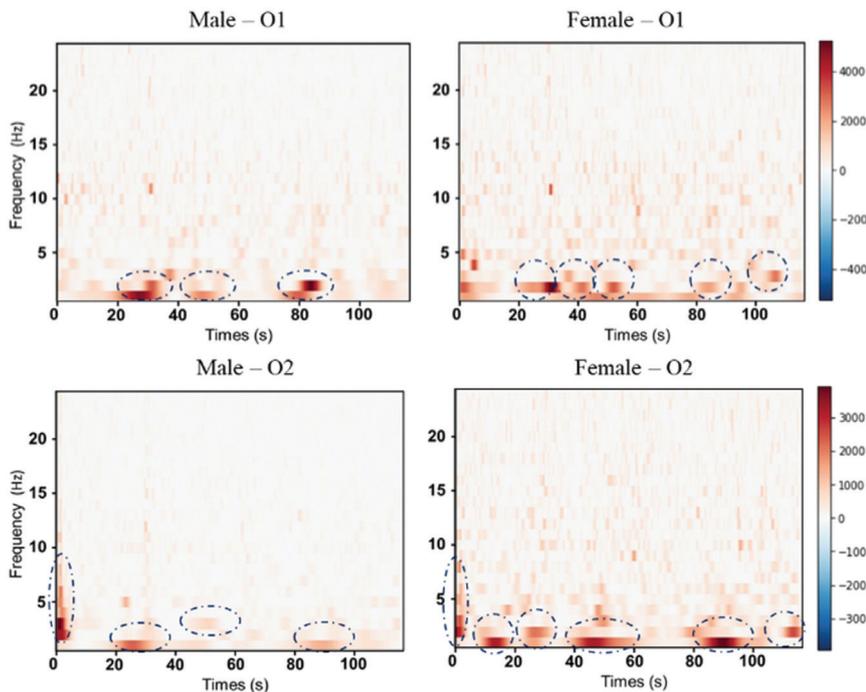


Fig. 10. O1 and O2 electrodes in males and females

In both male and female O1 and O2, there are numerous instances of high-frequency activities. Additionally, when actors appear, there are also occurrences of low-frequency activities. In this experiment, when a 2-second text appeared in a video, participants focused on reading it. As a result, high-frequency harmonics were observed on some channels at that moment. Additionally, female participants exhibited low-frequency activation each time a male actor appeared, a phenomenon also observed by male participants. Women may notice higher frequencies appearing more prominently in certain channels compared to men.

3. DEEP LEARNING METHOD FOR CLASSIFICATION

Before incorporating DL methods, we initially utilized the Power Spectral Density (PSD) analysis and the Common Spatial Pattern (CSP) algorithm in our research aimed at classifying gender based on EEG data. These methods were employed to extract pertinent features from EEG signals [52], which are crucial for accurate

gender classification. Since the features we are interested in are related to frequency, our approach involves analyzing the power spectrum density of EEG signals. This method allows us to examine the distribution of signal power across different frequency bands, which is essential for extracting relevant features for gender classification. The figure below illustrates the power spectral changes of the AF3, AF4, and T7 channels respectively for males and females.

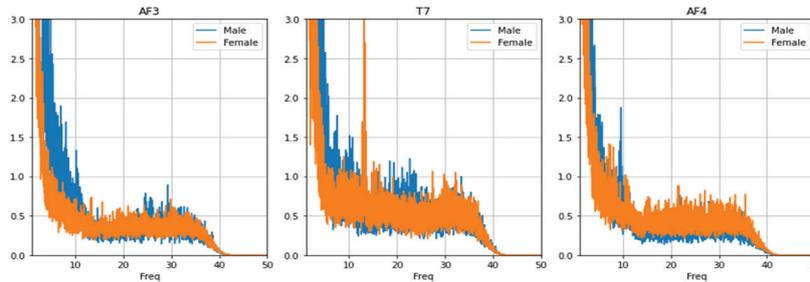


Fig. 11. AF3, AF4 and T7 channels PSD

First, we need to find a way to quantify the level of activity. We use the logarithm of variance of the signal within certain frequency bands as a feature for the classifier. The feature for the classifier will be the loga-

rithm of the variance of each channel. This will yield a single variable for each trial. The figure below shows a bar chart of the logarithm of variance for males and females.

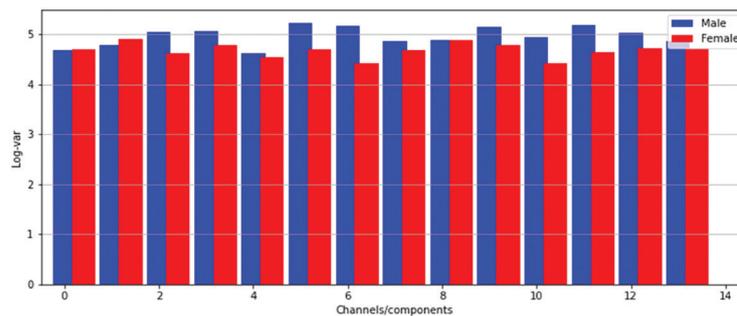


Fig. 12. Logarithm of variance for males and females

There are differences between the two classes that we can observe here. However, we need to maximize the difference between the male and female classes. We will use the CSP algorithm, which is designed to maximize the difference in variation between the two classes. We will find spatial filters that maximize the variance for one class and minimize the variance for the other class. The figure below shows the effect after applying this spatial filtering. Those are logarithm of variance features or components.

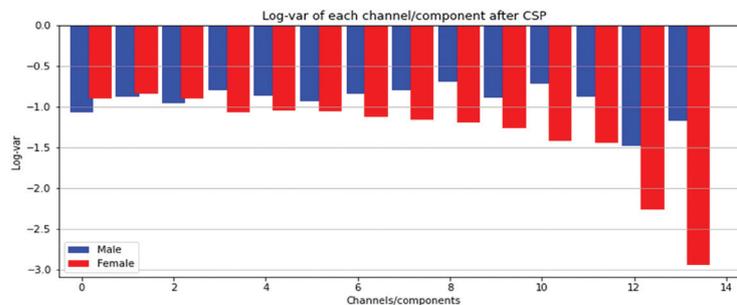


Fig. 13. Logarithm of variance for males and females after applying spatial filtering

These changes can be visualized as part of the power spectral density. The figure below visualizes the PSD af-

ter common spatial filtering. We can see a significant difference.

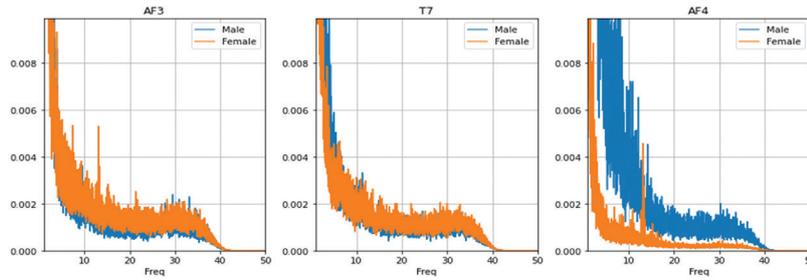


Fig. 14. PSD after common spatial filtering

We can now use these features to train a classifier and achieve good accuracy. We can discriminate between two classes as shown in the scatter plot on the two-dimensional plane below.

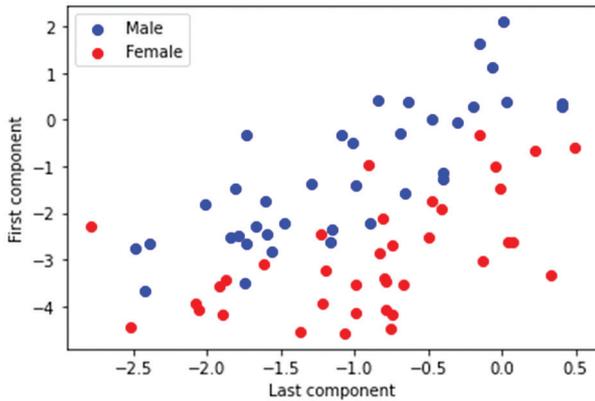


Fig. 15. CSP components between males and females

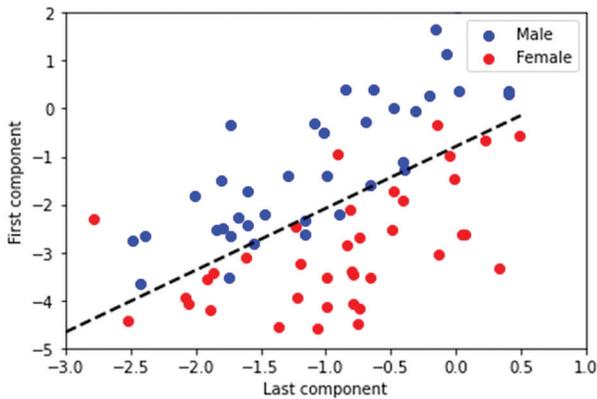


Fig. 16. Linear classification between two classes

Our decision boundary parameters have the following coefficients: $W: [2.57161 -1.9911]$, $b: 1.57859$. Using this method, we can classify males and females with 89.2% accuracy.

After that, we developed a DL model and tested it. Preprocessing was performed using a bandpass filter and we excluded 4 components using ICA basing on multiple tests and evaluations. The data was divided into 70% for training, 10% for validation, and 20% for testing. The figure below shows our developed model for gender classification.

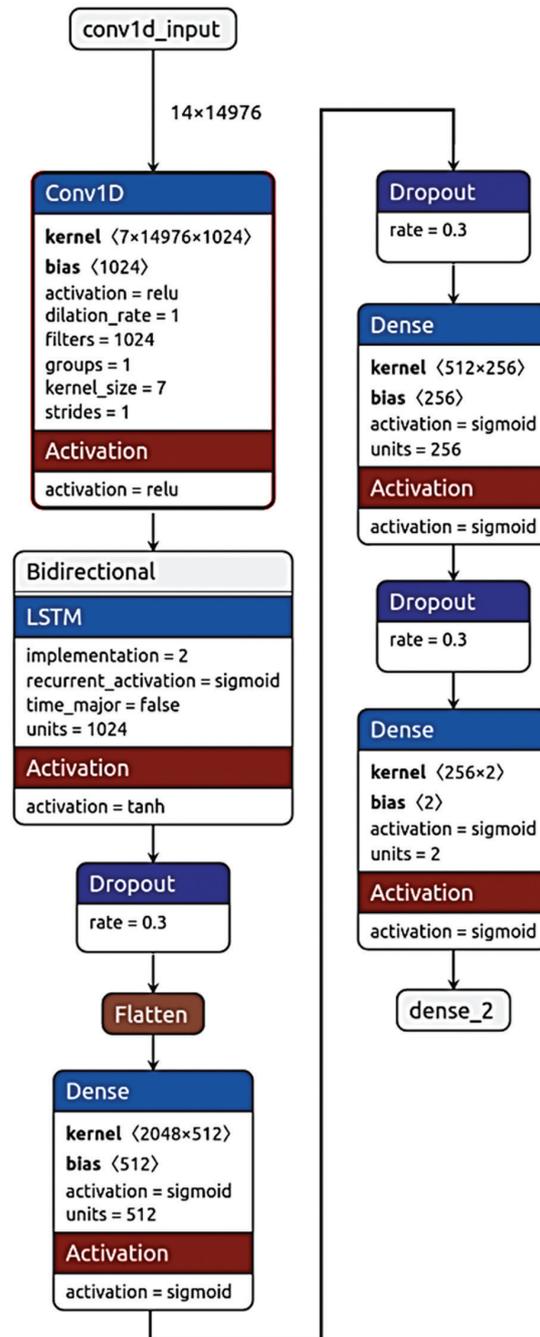


Fig. 17. Stream-based gender classification DL model

This model achieves 92% accuracy when using only preprocessed data.

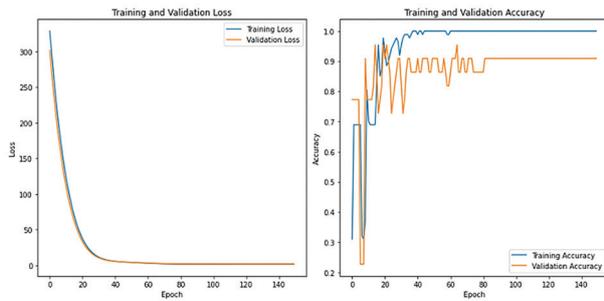


Fig. 18. Model results using preprocessed data

We added an additional step in data preparation. As previously discussed, the 'common pattern' of male/female EEG is subtracted from each participant's female/male EEG data alternately.

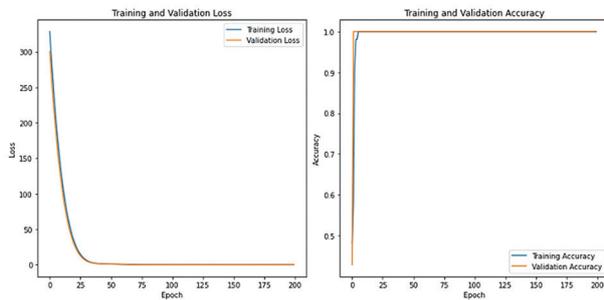


Fig. 19. Model results after average subtraction

With the assistance of this processed data, our DL model achieved remarkably high accuracy in accurately distinguishing between male and female genders. The data preprocessing, which involved subtracting the average EEG signal of females from each male participant's data and vice versa, significantly enhanced the model's ability to discern gender based on EEG patterns. In this case, using this processed data allows us to explore ways to make our model more efficient by potentially reducing its complexity. The confusion matrix is shown below the figure.

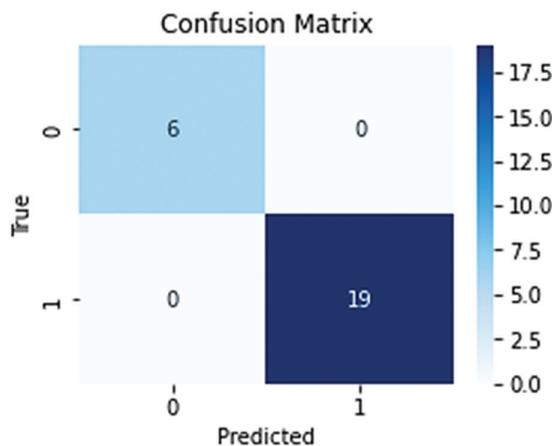


Fig. 20. Confusion matrix of our model

We correctly classified 6 females and 19 males based on our model's predictions.

4. CONCLUSION

This study presents a novel approach, the Stream-based Identification of Gender using Noninvasive Electroencephalographic Technology (SIGNET) method, by combining a deep learning model with spatiotemporal feature extraction for gender classification using EEG signals during video-streaming tasks, LSTM networks. Our method achieved 99% accuracy, highlighting its efficacy in distinguishing male and female brain activity during complex, dynamic tasks. These results demonstrate the potential of this framework in personalized medicine, neurotechnology, and gender-specific applications within AI.

Limitations of the method. While our study achieved promising results, several limitations must be acknowledged. The gender imbalance in our dataset (85 males and 37 females) presents challenges to the generalizability of our findings. Although equal sampling was ensured during training, future studies should aim for more balanced gender representation to validate these findings across a broader population. Additionally, our age distribution was concentrated between 17 and 23 years, which limits the generalizability of the results to older populations. Overfitting, due to the relatively small sample size, was a potential concern. To address this, we employed various strategies, including robust data preprocessing (e.g., ICA), gender-specific normalization and dropout regularization in the LSTM network to ensure generalization. These methods effectively mitigated overfitting and improved the model's robustness. Another limitation is the computational cost of deep learning models, particularly for real time applications. While the LSTM-based model demonstrated excellent performance in controlled environments, its scalability to real time systems or mobile platforms remains a challenge. Future research should focus on model optimization and edge computing to address this issue.

Key Findings. Our findings confirm that gender-based differences in EEG patterns are significant, particularly during tasks involving cognitive and emotional processes. Previous studies [33, 35] reported gender differences in cognitive tasks such as mental arithmetic and verbal memory. Similarly, our study highlights significant gender-related activity in the prefrontal cortex and occipital lobes regions critical for sensory processing and cognitive functions with more pronounced differences observed during the video-streaming task, which engaged both cognitive and emotional processing. Unlike traditional machine learning models, such as SVMs or Random Forests, which focus on static features, our use of LSTM networks allowed for the capture of temporal EEG patterns. This temporal sensitivity enabled the detection of subtle gender specific neural differences that static models might overlook. Furthermore, our application of wavelet transformations to analyze time-frequency features enhanced the model's ability to capture variations in frequency bands associated with gender differences.

Future Research and Application. Our future work aims to extend this study by incorporating larger and more diverse datasets to improve generalizability and explore additional factors such as age, cognitive state, and environmental influences that may affect EEG based gender classification. Addressing these factors will enhance the robustness and applicability of SIGNET in real world settings, opening new possibilities for personalized education systems and neuropsychological research. By leveraging deep learning models to analyze gender-specific EEG patterns, adaptive learning environments can be developed to cater to gender specific cognitive and emotional dynamics. Similarly, understanding gender differences in brain activity can contribute to more personalized therapeutic interventions for conditions such as ADHD and depression. In the realm of neurotechnology, this study has implications for improving the accuracy and security of biometric systems and BCIs by incorporating gender sensitive models. Future research can further explore how these findings might be extended to other demographic factors, such as age, cognitive states, and mental health conditions, potentially leading to more personalized and adaptive neurotechnology solutions. In addition, developing the SIGNET method for real time applications holds significant promise. The results of this study can be applied to personalized medicine, neurotechnology, AI applications, BCIs, and adaptive learning systems, providing a foundation for more advanced, gender-sensitive AI models designed for gender analysis.

5. ACKNOWLEDGMENT

This research was supported by "Capacity Building Project for School of Information and Communication Technology at Mongolian University of Science and Technology in Mongolia" (Contract No. P2019-00124) funded by KOICA (Korea International Cooperation Agency).

6. REFERENCES

- [1] A. Farkouh, T. Riedl, R. Gottardi, M. Czejka, A. Kautzky-Willer, "Sex-related differences in pharmacokinetics and pharmacodynamics of frequently prescribed drugs", *Advances in therapy*, Vol. 37, No. 2, 2020, pp. 644-655.
- [2] I. Adjabi, A. Ouahabi, A. Benzaoui, A. Taleb-Ahmed, "Past, present, and future of face recognition", *Electronics*, Vol. 9, No. 8, 2020, p. 1188.
- [3] S. E. Severiens, G. T. Ten Dam, "Gender differences in learning styles: A narrative review and quantitative meta-analysis", *Higher Education*, Vol. 27, No. 4, 1994, pp. 487-501.
- [4] N. Ellemers, "Gender stereotypes", *Annual Review of Psychology*, Vol. 69, No. 1, 2018, pp. 275-298.
- [5] K. Khan, M. Attique, I. Syed, A. Gul, "Automatic Gender Classification through Face Segmentation", *Symmetry*, Vol. 11, 2019, p. 770.
- [6] A. Swaminathan, M. Chaba, D. Sharma, Y. Chaba "Gender classification using facial embeddings: A novel approach", *Procedia Computer Science*, Vol. 167, 2020, pp. 2634-2642.
- [7] M. Shareef, T. Abd, Y. Mezaal, "Gender voice classification with huge accuracy rate", *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, Vol. 18, No. 5, 2020, pp. 2612-2617.
- [8] D. Kwasny, D. Hemmerling, "Gender and Age Estimation Methods Based on Speech Using Deep Neural Networks", *Sensors*, Vol. 21, No. 14, 2021, p. 4785.
- [9] C. T. Hsiao, C. Y. Lin, P. S. Wang, Y. T. Wu, "Application of Convolutional Neural Network for Fingerprint-Based Prediction of Gender, Finger Position, and Height", *Entropy*, Vol. 24, 2022, p. 475.
- [10] J. Xie, L. Zhang, J. You, D. Zhang, X. A. Qu, "Study of Hand Back Skin Texture Patterns for Personal Identification and Gender Classification", *Sensors*, Vol. 12, 2012, pp. 8691-8709.
- [11] A. Nunes, P. Serranho, P. Guimarães, J. Ferreira, M. Castelo-Branco, R. Bernardes, "When Sex Matters: Differences in the Central Nervous System as Imaged by OCT through the Retina", *Journal of Imaging*, Vol. 10, No.1, 2024, p. 6.
- [12] M. D'Angelo, L. Nanni, "Deep-Learning-Based Human Chromosome Classification: Data Augmentation and Ensemble", *Information*, Vol. 14, 2023, p. 389.
- [13] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, H. N. E. T. Mohamed, "Deep Iris: Deep Learning for Gender Classification Through Iris Patterns", *Acta Informatica Medica*, Vol. 27, No. 2, 2019, pp. 96-102.
- [14] M. A. R. Ahad, T. T. Ngo, A. D. Antar, M. Ahmed, T. Hossain, D. Muramatsu, Y. Makihara, S. Inoue, Y. Yagi, "Wearable Sensor-Based Gait Analysis for Age and Gender Estimation", *Sensors*, Vol. 20, 2020, p. 2424.
- [15] Z. Wang, J. Yang, Y. Zhu, "Review of Ear Biometrics", *Archives of Computational Methods in Engineer-*

- ing, Vol. 28, No.1, 2021, pp. 149-180.
- [16] A. N. Ruigrok, G. Salimi-Khorshidi, M. C. Lai, S. Baron-Cohen, M. V. Lombardo, R. J. Tait, J. Suckling, "Sex differences in the adult human brain: Evidence from 126 cases using VBM", *Neuroscience & Biobehavioral Reviews*, Vol. 39, 2014, pp. 34-50.
- [17] R. J. Haier, R. E. Jung, R.A. Yeo, K. Head, M. T. Alkire, "The neuroanatomy of general intelligence: sex matters", *NeuroImage*, Vol. 25, No. 1, 2005, pp. 320-327.
- [18] F. Sang, S. Zhao, Z. Li, Y. Yang, Y. Chen, Z. Zhang, "Cortical thickness reveals sex differences in verbal and visuospatial memory", *Cerebral Cortex*, Vol. 34, No. 3, 2024, p. 67.
- [19] J. N. Giedd, F. X. Castellanos, J. C. Rajapakse, A. C. Vaituzis, J. L. Rapoport, "Sexual dimorphism of the developing human brain", *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, Vol. 21, No. 8, 1997, pp. 1185-1201.
- [20] S. Genc, C. B. Malpas, G. Ball, "Age, sex, and puberty related development of the corpus callosum: a multi-technique diffusion MRI study", *Brain Structure and Function*, Vol. 223, 2018, pp. 2753-2765.
- [21] X. Wang, Y. Ren, Z. Luo, W. He, J. Hong, Y. Huang, "Deep learning-based EEG emotion recognition: Current trends and future perspectives", *Frontiers in Psychology*, Vol. 27, No. 14, 2023, p. 1126994.
- [22] G. Yoo, H. Kim, S. Hong, "Prediction of cognitive load from electroencephalography signals using long short-term memory network", *Bioengineering*, Vol. 15, No. 10, 2023, p. 361.
- [23] N. Jirakittayakorn, Y. Wongsawat, S. Mitirattanakul, "ZleepAnlystNet: a novel deep learning model for automatic sleep stage scoring based on single-channel raw EEG data using separating training", *Scientific Reports*, Vol. 14, No. 1, 2024, p. 9859.
- [24] S. J. Shah, A. Albishri, S. S. Kang, Y. Lee, S. R. Sponheim, M. Shim, "ETSNet: A deep neural network for EEG-based temporal-spatial pattern recognition in psychiatric disorder and emotional distress classification", *Computers in Biology and Medicine*, Vol. 1, 2023, p. 106857.
- [25] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals", *Neural Computing and Applications*, Vol. 35, No. 20, 2023, pp. 14681-14722.
- [26] F. Wang, Y. Wan, M. Li, H. Huang, L. Li, X. Hou, J. Pan, Z. Wen, J. Li, "Recent Advances in Fatigue Detection Algorithm Based on EEG", *Intelligent Automation & Soft Computing*, Vol. 35, No. 3, 2023.
- [27] S. Poorani, P. Balasubramanie, "Deep learning based epileptic seizure detection with EEG data", *International Journal of System Assurance Engineering and Management*, Vol. 2, No. 1, 2023, pp. 1-10.
- [28] M. M. Abdulghani, W. L. Walters, K. H. Abed, "Imagined Speech Classification Using EEG and Deep Learning", *Bioengineering*, Vol. 10, No. 6, 2023, p. 649.
- [29] N. Marvi, J. Haddadnia, M. R. Bordbar, "An automated drug dependence detection system based on EEG", *Computers in Biology and Medicine*, Vol. 158, 2023, p. 106853.
- [30] N. G. Seshadri, S. Agrawal, B. K. Singh, B. Geethanjali, V. Mahesh, R. B. Pachori, "EEG based classification of children with learning disabilities using shallow and deep neural network", *Biomedical Signal Processing and Control*, Vol. 82, 2023, p. 104553.
- [31] C. M. Segning, R. A. Silva, S. Ngomo, "An Innovative EEG-Based Pain Identification and Quantification: A Pilot Study", *Sensors*, Vol. 12, 2024, p. 3873.
- [32] Y. Wada, Y. Takizawa, Y. Zheng-Yan, N. Yamaguchi, "Gender differences in quantitative EEG at rest and during photic stimulation in normal young adults", *Clinical Electroencephalography*, Vol. 25, No. 2, 1994, pp. 81-85.
- [33] W. Skrandies, P. Reik, C. Kunze, "Topography of evoked brain activity during mental arithmetic and language tasks: sex differences", *Neuropsychologia*, Vol. 37, No. 1, 1999, pp. 421-430.
- [34] B. Rescher, P. Rappelsberger, "Gender dependent EEG-changes during a mental rotation task", *International Journal of Psychophysiology*, Vol. 33, No. 3, 1999, pp. 209-222.

- [35] N. V. Volf, O. M. Razumnikova, "Sex differences in EEG coherence during a verbal memory task in normal adults", *International Journal of Psychophysiology*, Vol. 32, No. 2, 1999, pp. 113-122.
- [36] N. Jaušovec, K. Jaušovec, "Sex differences in brain activity related to general and emotional intelligence", *Brain and Cognition*, Vol. 59, No. 3, 2005, pp. 277-286.
- [37] M. J. Larson, M. South, P. E. Clayson, "Sex differences in error-related performance monitoring", *Neuroreport*, Vol. 22, No. 1, 2011, pp. 44-48.
- [38] S. E. Kober, C. Neuper, "Sex differences in human EEG theta oscillations during spatial navigation in virtual reality", *International Journal of Psychophysiology*, Vol. 79, No. 3, 2011, pp. 347-355.
- [39] J. Langrova, J. Kremláček, M. Kuba, Z. Kubova, J. Szanyi, "Gender impact on electrophysiological activity of the brain", *Physiological Research*, Vol. 61, 2012, p. S119-S127.
- [40] P. Nguyen, D. Tran, T. Vo, X. Huang, W. Ma, D. Phung, "EEG-based age and gender recognition using tensor decomposition and speech features", *Proceedings of Neural Information Processing: 20th International Conference, Daegu, Korea, 3-7 November 2013*, pp. 632-639.
- [41] P. Kaushik, A. Gupta, P. P. Roy, "EEG-based Age and Gender Prediction Using Deep BLSTM-LSTM Network Model", *IEEE Sensors Journal*, Vol. 19, No. 7, 2018, pp. 2634-2641.
- [42] P. Wang, J. Hu, "A hybrid model for EEG-based gender recognition", *Cognitive Neurodynamics*, Vol. 13, No. 6, 2019, pp. 541-554.
- [43] B. Kaur, D. Singh, P. P. Roy, "Age and gender classification using brain-computer interface", *Neural Computing and Applications*, Vol. 31, No. 10, 2019, pp. 5887-5900.
- [44] J. Zhang, H. Xu, L. Zhu, W. Kong, Z. Ma, "Gender recognition in emotion perception using EEG features", *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, San Diego, CA, USA, 18-21 November 2019*, pp. 2883-2887.
- [45] B. Bučková, M. Brunovský, M. Bareš, J. Hlinka, "Predicting sex from EEG: validity and generalizability of deep-learning-based interpretable classifier", *Frontiers in Neuroscience*, Vol. 14, 2020, p. 589303.
- [46] A. Markovic, M. Kaess, L. Tarokh, "Gender differences in adolescent sleep neurophysiology: a high-density sleep EEG study", *Scientific Reports*, Vol. 10, No. 1, 2020, p. 15935.
- [47] N. K. Al-Qazzaz, M. K. Sabir, S. H. Ali, S. A. Ahmad, K. Grammer, "Multichannel optimization with hybrid spectral-entropy markers for gender identification enhancement of emotional-based EEGs", *IEEE Access*, Vol. 9, 2021, pp. 107059-107078.
- [48] A. E. Cave, R. J. Barry, "Sex differences in resting EEG in healthy young adults", *International Journal of Psychophysiology*, Vol. 161, 2021, pp. 35-43.
- [49] A. Hassani, A. Hekmatmanesh, A. M. Nasrabadi, "Gender differences in EEG responses to color and black-white images: implications for neuromarketing strategies", *IEEE Access*, Vol. 11, 2023, pp. 93739-93753.
- [50] T. Jochmann, M. S. Seibel, E. Jochmann, S. Khan, M. S. Hämäläinen, J. Haueisen, "Sex-related patterns in the electroencephalogram and their relevance in machine learning classifiers", *Human Brain Mapping*, Vol. 44, No. 14, 2023, pp. 4848-4858.
- [51] Y. Niu, X. Chen, Y. Chen, Z. Yao, X. Chen, Z. Liu, X. Meng, Y. Liu, Z. Zhao, H. Fan, "A gender recognition method based on EEG microstates", *Computers in Biology and Medicine*, Vol. 173, 2024, p. 1366.
- [52] B. Blankertz, G. Dornhege, M. Krauledat, K. R. Müller, G. Curio, "The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects", *NeuroImage*, Vol. 37, No. 2, 2007, pp. 539-550.

LTE Coverage Planning Based on Improved Grey Wolf Optimization

Original Scientific Paper

Fekar Mohammed Riyadh El Mansour*

University of science and technology of Oran
Faculty of Electrical Engineering, Signals and Images Laboratory, Department of Electronics
Oran, Algeria
fekarmohammedriyadhelmansour@gmail.com

Guezouri Mustapha

University of Oran1-Ahmed Benbella, Faculty of exact and applied sciences
Department of Computer Science
Oran, Algeria
guezouri.mustapha@univ-oran1.dz

*Corresponding author

Abstract – Automatic planning and dimension optimization of LTE is one of the crucial tasks in the mobile networking community. It is well known that this process is an NP-hard issue that requires huge computing resources. We also noticed that the actual proposed solutions are still inefficient in terms of scalability (handling a large number of eNodeBs) and runtime effectiveness. Moreover, SINR handling and variability of propagation loss models with respect to areas' types further complicate the coverage planning task. In this paper, we propose a swarm intelligence-based method for effectively placing and configuring the eNodeBs of an LTE network. In particular, we propose two variants of grey wolf optimizer (GWO), namely a discrete version of GWO (DGWO) and an improved version of GWO (IGWO) for LTE coverage planning. The improved version consists of an additional local search rule that allows for exploring regions closer to the promising solutions. The approaches are simulated on an urban area with many types of clutter. The IGWO technique had a coverage of 99.0% of 10 dB SINR rate and 95.1% of 12 dB SINR rate. The obtained results show that IGWO is more effective than the discrete one and other existing metaheuristics in terms of cost and coverage rates. More specifically, it ensures a coverage improvement (with respect to 10 dB SINR rate) of 10.6%, 10.5%, and 2.6 % in comparison to DGWO, Tabu search (TS), and discrete particle swarm optimization (DPSO) respectively.

Keywords: cellular planning, coverage planning, grey wolf optimizer, metaheuristics, LTE

Received: September 27, 2024; Received in revised form: January 17, 2025; Accepted: January 20, 2025

1. INTRODUCTION

Cellular radio networks recognize a high evolution speed, which is mainly driven by the user's needs, such as reliable coverage, traffic capacity, and other QoS metrics. To meet these requirements, Long Term Evolution (LTE) networks provide a set of ingredients to ensure reliable communications, user's throughput satisfaction, low latency (for some specific applications), and better scheduling schemes of radio resources [1, 2]. Regarding the increase in the communication data rate, LTE networks must face the delay spread caused by multipath propagation of the transmitted signal (which increases the duration of the symbol by up to 1-5 micro

seconds, causing the interference with the subsequent symbol). Consequently, this delay spread gets translated into frequency-selective fading, which deteriorates the SINR (Signal-to-Interference-plus-Noise Ratio) and hampers the data rate improvement. To palliate this issue, the LTE standard uses multiple subcarriers of lesser bandwidth (typically 15 KHz) which are orthogonal and more resilient to frequency-selective fading channels (this division technique is called orthogonal frequency division multiplexing or OFDM). As a result, the inter-carrier interference is significantly reduced, and the total spectrum is efficiently managed. Furthermore, the LTE standard adopts a flexible set of frequency bands

with different bandwidths (from 1.4 MHz to 20 MHz) that serve the needs of regions with varying user densities and varying data rate requirements. We mention that the extension of OFDM for serving different users at the same time, also known as orthogonal frequency division multiple access, is used in the downlink stream of LTE; however, the single carrier frequency division multiple access (SC-FDMA) is used in the uplink stream because of the low processing capabilities of the user equipment (UE). It is worth highlighting that the issue of high peak-to-average-power ratio can be palliated with SC-FDMA, and therefore complex operations are not needed in UE. For the same purpose of enhancing the data rates and channel reliability, the use of multiple-input-multiple-output systems (MIMO) allows for reinforcing the diversity (and therefore the enhancement of the overall SINR that is targeted in the coverage planning) and spatial multiplexing (and accordingly an improvement in capacity). Furthermore, the availability of sophisticated computing power in eNodeBs (and sometimes UE) enables the execution of crucial operations such as inverse fast Fourier transforms (IFFT) and matrix processing (such as precoding and beamforming in MIMO settings). In sum, these enabling capabilities (MIMO, OFDMA) allow the fulfillment of a high spectral efficiency of the network, which may reach 15 bits/Hz for the 20 MHz bandwidth.

Planning an LTE network is a complicated task that may involve several potentially conflicting objectives concerning coverage, capacity, cost, power, and eventually other metrics [3, 4]. Further, the lack of an accurate propagation model that completely involves all the geographical details of the studied area (e.g., terrain, vegetation, building height, etc.) will hamper the attainment of a satisfying coverage result. Moreover, the temporal variability of the abovementioned geographical factors further exacerbates the quantification of the shadowing and interference (and this situation will negatively impact the coverage planning issue). It is estimated that cellular radio network operators need to extend the capacity for dozens of billion connections in the future [5]. Moreover, the theoretical complexity of network planning is known to be NP-hard [6, 7], and consequently, manual simulations or ad-hoc planning methods are not adequate for large-scale problems.

According to [8], network planning is the process of estimating, placing, and configuring a set of base stations (BSs¹, or eNodeBs in the LTE context) to ensure the coverage and capacity of a given area. There are three steps in network planning: pre-planning, also called dimensioning; detailed planning; and post-planning or optimization [8, 9]. The dimensioning phase consists of roughly estimating the number of BSs needed to cover an area of interest. This result is considered as an input of the detailed-planning phase. The detailed planning consists of deciding the physical locations of BSs in the

target area of interest. Finally, the optimization step is a post-deployment task that consists of rectifying the network performance after having analyzed the ground measurements.

Coverage planning is a crucial phase in the deployment of a cellular network. It involves choosing an adequate propagation model based on the area's terrain, clutter, and population characteristics [10]; in addition, coverage planning is aimed at satisfying the constraint of having a received signal power greater than a predefined threshold in every location of the target area. Sometimes, this coverage constraint is aimed at having a SINR level greater than a predefined threshold at every point of the studied area. As stated before, coverage planning is an NP-complete problem [7, 11], and accordingly, there is no efficient (polynomial) algorithm that can solve it with a perfect optimality rate.

In general, the existing works on cellular planning mainly focus on a reduced set of conflicting objectives (such as cost, coverage, capacity, power consumption, and handover zone management), but there have been few initiatives to tackle the fully automatic placement and configuration of BS [9, 12]. Moreover, there is a lack of approaches that leverage artificial intelligence (AI) and swarm intelligence (SI) in optimizing this task. In fact, numerous works adopt simplistic heuristics or semi-automatic methods for estimating or placing BS [13, 14]. However, totally automatic AI-based methods are still sparsely available in the field of LTE planning. Moreover, the comprehensive analysis of the impact of SI's hyperparameters on the planning quality (e.g. SINR levels) are still insignificant the literature. To handle this gap, we propose in this paper an improved version of a Grey Wolf Optimizer (GWO) [15], termed IGWO, that addresses the automatic placement and configuration of eNodeBs so as to best meet the coverage and cost requirements of an LTE network. It is worth noting that swarm-intelligence-based methods (and specifically GWO variants) can be used to handle larger sizes of search space problems, reduce the time taken to achieve near-optimal planning configurations, and ensure high agility to the changes observed in the coverage or capacity of the studied network. Moreover, the use of a variant based on GWO can spark a significant improvement in the quality of the retained solutions. In fact, the use of a majority voting rule and a set of guiding agents in the search (instead of a single agent, as is the case for other swarm intelligence-based methods), may improve the quality of the retained optimum under some assumptions [16]. Furthermore, adjustment of the perturbation distance can help orient the search for more promising regions and avoid less attracting parts of the cost function landscape. Inspired from the voting theory, we advocate that the more the diversity of guiding agents is high, the better the quality of the retained solutions.

The key contributions of this paper can be summarized as follows:

¹ <https://www.forsk.com/atoll-overview>

- The optimal eNodeB placement/configuration of an LTE network is implemented using a discretized GWO (DGWO) algorithm that optimizes the deployment cost and coverage; furthermore, the adjustment of the SINR thresholds will directly enhance the network capacity.
- An improved version of the previous GWO, termed IGWO, is designed and applied in the studied area, with the addition of a new operator based on a local search. This latter one replaces the perturbed distance move used in the standard GWO and aims to bring an added value to the search. A detailed comparison among DGWO, IGWO, and other metaheuristics (e.g., TS, PSO) is demonstrated in the evaluation section.
- The impact of different population sizes on the cost of eNodeB deployment, coverage rate, and CPU time is evaluated.
- The remainder of this paper is organized as follows: Section 2 contains a review of the existing works on LTE planning. In Section 3, we formulate the problem as a multi-objective optimization issue with constraints. Section IV presents the GWO and IGWO optimization algorithms. Section V presents the results and the related discussions. Finally, Section 6 specifies the conclusion and future directions.

2. LITERATURE REVIEW

Many works have been proposed recently in the field of cellular radio planning [12, 17- 20]. They mainly differ in terms of the planning algorithm (simple heuristics, exponential methods, metaheuristics), the optimized goals (pertaining to coverage, capacity, cost, handover management, and power consumption), and the leveraged input/output parameters. For instance, some papers specify site locations, traffic models, BS configurations, propagation models, type of base station, and frequency reuse strategy [18-20]. Other works, such as [17], investigate all the alternatives that can be used to enhance the coverage for both Long Term Evolution (LTE) and 5G mobile networks. These alternatives include network deployments, frequency bands, and interference mitigation. Along the same lines, Elsayw et al. [20] propose a rigorous mathematical model based on stochastic geometry [21] for analyzing the coverage of cellular radio networks; specifically, the authors used random networks to approximate the SINR score and accordingly other related metrics such as outage probability and average data rate.

In what follows, we cover the main categories for achieving the best coverage and capacity in the cellular planning.

2.1. SIMPLE HEURISTIC/SIMULATION-BASED METHODS

The works in this category address only a part of the whole problem or exploit properties (or heuristics) to

reduce the problem complexity. For example, the work cited in [22] addresses partial aspects (such as azimuth tuning) of the cellular planning using the divide-and-conquer strategy.

In [23], the authors exploit the maximum allowable path loss (MAPL) heuristic to estimate the cell area and accordingly predict the dimensioning of the network. Thereafter, a comprehensive set of simulations is conducted to determine the BS locations.

In [24], the authors addressed both coverage and capacity planning; the final number of estimated BSs is the maximum given by the procedures that resolve each of them. To solve the coverage/capacity planning, the proposition leverages empirical models and statistical formulas to estimate the cell area, number of users, data volume, and user's throughput. The authors do not handle the location of BS sites.

The work presented in [9] addresses LTE dimensioning by leveraging three scenarios: macrocell deployment, small cell deployment, and heterogenous deployment. The BS deployment is deemed acceptable if it meets the requirements of all users (i.e., the satisfaction of both target uplink throughput and target downlink throughput).

The adopted algorithm starts from a superfluous number of initially deployed BSs. Then, it gradually removes the redundant BSs until it arrives at near-optimal set of BSs (from which any site reduction will cause a dramatic deterioration of the throughput metric of the cellular network).

The study presented in [25] leveraged the Cost-Hata propagation model in order to set the locations of LTE BSs. The conducted simulations assessed several planning parameters, such as Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), SINR, and throughput.

In the context of 5G-NR planning, the authors in [26] analyzed and compared a plethora of propagation models (such as Knife Edge Diffraction [KED] and Dominant Path model [DPM], 3rd Generation Partnership Project [3GPP] and ASTER) for both the millimeter-wave band (28GHz) and the 3.5 GHz band. The work targeted an urban area (Quito city) in Ecuador and leveraged multiple configurations of MIMO (2×2 and 4×4 , 64×64) streams. More specifically, the planning is performed using the platform of Atoll, and it is evaluated using the metrics of throughput, RSRP, and RSRQ indicators; the analysis also provides consistent insights to choose a suitable propagation model for deploying a 5G-NR network.

To achieve the 5G NR planning, the authors used the C-band/mm-wave bands. More specifically, they considered the maximum allowable path loss (MAPL) heuristic and several propagation models (which are compatible with 5G), namely, Urban Macro model (Uma) and Urban Micro model for both line-of-sight (LOS) and no-line-

of-sight (NLOS) cases to derive the cell radius (of the gNodeB). Then, the coverage planning was performed using Atoll environment and the estimated number of gNodeBs; simulations were achieved for two operating frequencies: 3.5 GHz and 28 GHz. The authors derived the levels of RSRP and SNIR and downlink throughputs using the Atoll simulations over the urban area.

To handle the poor coverage issue of eNodeBs, the authors in [27] leveraged the automatic cell planning (ACP) optimization module of Atoll software to improve the LTE network planning in the city of Solok (Indonesia). In particular, the authors enhanced the RSRP and SINR parameters. This optimization method enables deriving the best setting of sectoral antenna parameters (tilting, azimuth, and antenna height) and allows for palliating the coverage holes and the interferences issues.

2.2. APPROXIMATION-BASED METHODS

In such methods, the optimality of the solution is compromised with the efficiency of the search algorithm. Particularly, these approaches prefer approximate (and efficient) methods to get near-optimal solutions. The works presented in [28, 29] exploit the division of global planning problems into local ones using heuristics of locality. Then, the solutions of individual local problems are fused to derive approximate global solutions.

2.3. EXHAUSTIVE/EXPONENTIAL METHODS

These methods exploit exhaustive search or exponential methods coupled with heuristics to alleviate the exploration overhead of BS assignment/configuration. The works presented in [30, 31] detail concrete implementations of exhaustive search methods for cellular planning.

2.4. METAHEURISTIC-BASED METHODS

In this category, the authors leverage higher-level optimization algorithms, such as particle swarm intelligence (PSO) [13], genetic algorithms [5, 12, 32, 33], and bee colony optimization [34], to efficiently solve NP-hard planning problems.

The authors of [12] simultaneously address the objectives of coverage, capacity, and power consumption in LTE planning. The coverage sub-goal is met by satisfying a minimum value of the received signal strength at each position of the target area. Moreover, each eNodeB is characterized by its position, tilt, sector, azimuth, and transmit power.

To ensure efficient eNodeB placement and configuration, the authors used the multi-objective metaheuristic of nondominated sorting genetic algorithm-II (NSGA-II). This algorithm is able to reach the pareto optimal solutions (evaluated in terms of the previous three objective functions) in an adequate time.

A genetic algorithm is proposed in [7] to plan an LTE network with mixed cell structures (including macro

cells, micro cells, and relay nodes). The model addresses both coverage and capacity while minimizing the cost of cells.

In [8], the researchers leveraged simulated annealing (SA) to handle the dimensioning and placement of BSs in LTE networks for optimal capacity satisfaction; they handled both macrocell deployment and hybrid macro and microcell deployment.

The SA metaheuristic involves conducting a random search with perturbations on the parameters of the utility function (which consists of the satisfaction of users' throughputs). We also highlight that the algorithm occasionally accepts moves that increase the scores of the objective function with a small probability, p , that is inversely proportional to the temperature hyperparameter of SA.

In [35], the authors proposed a hybrid dragonfly algorithm [36] with differential evolution (DADE) for LTE cell planning in vulnerable areas or post-disaster zones. The proposition takes into account the user coverage, user association, and capacity requirements. Moreover, the emergent dragonfly metaheuristic leverages a set of reflexes such as cohesion, food source attraction, alignment, separation and distraction from enemies to improve the advances in the search space. These five reflexes, in addition to the momentum term, are combined to compute the new position of the agent (i.e., dragonfly); the results indicate a high performance with respect to other metaheuristics such as PSO, differential evolution (DE), whale optimization algorithm (WOA), and GWO.

The authors of [13] tackled both coverage and capacity planning of LTE networks. First, they deduced the approximate number of BSs using MAPL and the estimated cell surface. Then, they applied two metaheuristics to determine the best locations of all BSs that would allow maximizing both coverage and capacity: PSO and GWO. The PSO/GWO simulation was performed on an area of 100 km² with a carrier frequency of 1.8 GHz and 10 MHz bandwidth. The minimum data rate used in the capacity optimization (for every user) was set to 64 kbps for the uplink channel and 1 Mbps for the downlink channel.

In [33], the authors presented an Adaptive Variable Length Genetic Algorithm (AVLGA) as a novel metaheuristic for optimizing BS positioning in LTE networks. The proposed approach employs a weighted fitness function that combines the coverage and capacity.

The major enhancement of this proposition is the acceptance of length-variable solutions in genetic algorithms (GA). The results show a better management of interferences with respect to the conventional optimization techniques.

In [37], the authors optimized the BS locations in order to plan the coverage and capacity of LTE networks (the objectives were combined with a nonlinear scheme).

They implemented the standard GWO metaheuristic and a GWO variant to handle the optimization of BS locations. The variant leveraged two patterns for varying a hyperparameter that controlled the exploration/exploitation trade-off. The results confirmed the superiority of the GWO variant with respect to the standard one.

In [38], the authors leveraged GA to optimize the coverage probability of LTE networks (which is based on the SINR computation). They used three decision variables (BS location, BS height, and BS transmission power) to tune the deployment of BSs.

The authors of [39] used the Tabu search metaheuristic to plan 3G UMTS networks. More specifically, they aimed at tuning the BS configuration (tilt, azimuth, power allocation) and placement in order to enhance the coverage and capacity objectives.

In [40] the authors addressed the planning of 5G BSs in the urban area of Thapathali (Nepal), using both the frequency carriers of 28 GHz and 3.6 GHz (the first scenario includes the mmWave frequencies). Different metaheuristics were employed to determine the optimal placement of 5G BSs that would maximize the coverage and capacity, minimize the interference, and improve the cost. These alternatives include GA, PSO, SA, and GWO. All these techniques involve exploring the possible BS configurations and locations to derive the best trade-off. Before applying the metaheuristics, the authors estimated the radius using link budget analysis and wireless propagation models that are compatible with 5G.

The obtained results indicate that PSO showed superior performance in all the metrics (including the coverage, capacity, and cost of infrastructure) and for both the 28 GHz and 3.6 GHz operating frequencies; GA also provided a satisfiable performance, ranking second in terms of the overall performance.

The works by [41] is focused on the tuning of resource block (RB) distribution and power allocation in the context of optimizing both the coverage and capacity of LTE networks. To this end, the authors employed two techniques: GA and the water filling algorithm. Both algorithms allow for finding the near-optimal assignment of resource blocks to the sub-cells of the network.

In [42], the authors leveraged genetic algorithms to optimize the antenna positioning problem (APP) and enhance the coverage planning of LTE networks. To meet this objective, the authors implemented seven empirical models for propagation losses (including, Hata and Cost-231-Hata). The results showed that an appropriate empirical model selection is able to derive a satisfying performance of PPA for all area types (urban, suburban, and rural areas).

2.5. MACHINE LEARNING/ HYBRID METHODS

The overwhelming complexity of cellular network planning can be palliated using machine learning [43] or hybrid approaches [44] that may combine heuristics,

metaheuristics, and machine learning. Following this line of thought, the authors of [43] leveraged both reinforcement learning (more specifically, the actor-critic method) and Bayesian optimization to select the best parameters for BSs (transmit power and tilt). The two objective functions handled by the authors are the under-coverage (which handles the coverage holes) and the over-coverage (which handles the interferences emanating from other cells). Under-coverage means that the maximum signal power received (such as RSRP in LTE) from the cell sector antenna is larger than a given threshold T_1 , and over-coverage means that the difference between the maximum received signal power (e.g., RSRP) and the sum of received powers from other cells does not exceed a second threshold T_2 (common values for T_1 and T_2 are -110 dBm and 6 dB).

In this setting, the aim is to identify the set of Pareto optimal solutions that best balance the two objectives. The reinforcement learning model contains two neural networks; the first one is a deep neural network (called actor) that learns the policy using gradient decent and predicts as outputs the configuration parameters of the sector antennas. The second network (called critic) learns the Q-score of each pair (state, action), (which is also seen as the long-term reward).

In [45], the authors addressed the planning and optimization of LTE networks. The coverage planning is accomplished through randomly segmenting the possible locations of BSs into groups; this segmentation is done with respect to the Channel Quality Index (CQI) heuristic. Thereafter, an exhaustive search is executed in each segment to derive the near-optimal solution.

In contrast to the majority of cellular planning works (that are concerned with lowering the number of BSs), the authors of [44] address network planning by modeling it as a power consumption minimization problem with cell coverage rate and cell load constraints. This energy-oriented approach is based on fuzzy c means clustering to choose the BS locations.

In [46], the authors provided a hybrid approach involving three stages. In the first phase, a feasible solution to the issue is calculated using constraint satisfaction strategies embedded with a tailored heuristic search. The second phase consists of conducting a local search to improve the found solution. The third phase entails further enhancing to the solution calculated through the second phase.

In [47], the authors highlighted the lack of accuracy in path loss empirical models, especially in some physical area scenarios (such as flat areas found in plains and other rural areas). To achieve accurate planning of LTE networks, the authors proposed a neural network model combined with the PSO technique to correct the standard propagation model (SPM). This latter adjusted SPM model was applied to palliate the planning problems such as over-coverage, coverage holes, overlapping coverage, and high interference ratios.

Table 1 the main approaches designed for network planning. For the sake of conciseness, we denote the cost as *co*, capacity as *ca*, coverage as *cov*, and power consumption as *po*.

Table 1. Main classes of cellular planning approaches

Approach	Technique Used for BS Placement	Supported Objective Functions	Target Networks
[21]	Heuristics and simulations	co, cov	LTE
[23]		co, cov	
[9]		ca, po, co	
[48]		cov, ca	
[7]	GA	co, ca, cov,	LTE
[33]		cov, ca, co	LTE
[38]		po, cov, co	LTE
[14]		po, ca	5G
[12]		ca, po, cov	LTE
[13]	GWO	ca, cov	LTE
[37]		ca, cov	
[34]	Bee colony optimization	cov, co, po	WIMAX
[39]	TS	co, ca, cov, po	3G UMTS
[13]	PSO	ca, cov	LTE
[40]		ca, cov, co	5G
[8]	SA	ca, co	LTE
[40]		ca, cov, co	5G
[43]	Machine learning	cov (with balancing)	LTE
[44]		ca, co, po	
[45]	Hybrid approach	co, ca, cov	LTE
[46]	Hybrid approach	co, Handover	3G UMTS

3. PROBLEM STATEMENT

As stated earlier, the cellular network planning problem is known to be an NP-complete problem, and in fact, it can be considered as an instance of the set covering problem (SCP) [49]. Before presenting its mathematical expression, we first introduce some necessary concepts that are crucial for the subsequent parts.

The channel gain (in dB scale) between the user equipment UE located at the point *k* and the BS *j* over the sub-band (or the subcarrier) *i* is given by [9]:

$$H_{kij} = (-C - \alpha \log_{10}(d_{kij})) - \xi_{kij} + 10 \log_{10} F_{kij} \quad (1)$$

The first term indicates propagation loss, with *C* representing the path loss constant, d_{kij} the distance in km from the user's location, *k*, to BS *j*, and α the path loss exponent. The second term, ξ_{kij} corresponds to log-normal shadowing, with zero mean and a standard deviation σ . The last factor, F_{kij} stands for the fast fading power (or small-scale fading). F_{kij} follows the Rayleigh distribution. Since cellular network planning is mainly concerned with large-scale fading, we only consider the first and second terms (path propagation loss and shadowing) in the SINR definition (see equations (13) and (16)). This large-scale fading (the combination of propagation loss and shadowing) is denoted as LSF_{kij} .

$$LSF_{kij} = (-C - \alpha \log_{10}(d_{kij})) - \xi_{kij} \quad (2)$$

The path propagation loss can be calculated using one of the well-known empirical models, such as Cost231-Hata model. This latter model is focused on the frequency range of 1500 MHz up to 2000 MHz and estimates the loss as follows [50]:

$$PL = L + C_m \quad (3)$$

$$L = 46.3 + 33.9 \times \log(f) - 13.92 \log(H_{bs}) - a(H_r) + (44.9 - 6.55 \times \log(H_{bs})) \times \log(d) \quad (4)$$

$$a(H_r) = (1.1 \log(f) - 0.7) H_r - (1.56 \log(f) - 0.8) \quad (5)$$

Where:

H_{bs} : The BS height (m).

H_r : The receiver antenna height (m).

d: The distance between the BS and MS (m).

f: The carrier frequency (MHz).

C_m : Constant factor (in urban zones, it is 3 dB).

Before defining the objective function of the planning, we introduce the following parameters:

M is the total number of all possible eNodeBs (BS).

N is the total number of possible locations in the studied

$$\text{area: } N = \frac{\text{AreaLength}}{\text{ResolutionStep}} \times \frac{\text{AreaWidth}}{\text{ResolutionStep}} \quad (6)$$

The length, width, and ResolutionStep are defined in Section 5. In the following, we assume that each location point in the studied area is defined by the latitude and longitude coordinates (*x*, *y*).

If we denote the locations/configurations ($x_1, y_1, \text{config}(1)$), ..., ($x_M, y_M, \text{config}(M)$) of the *M* possible eNodeBs as Sol, then the problem of coverage planning can be defined as follows:

$$\text{Min TotalCost(Sol)} = \sum_{j=1}^M \text{Position}_{j,x_j,y_j} * \text{Cost}_{eNB}(x_j, y_j, \text{config}(j)) \quad (7)$$

Where,

$$\sum_{\text{all}(x,y) \in \text{Area}} \text{Position}_{j,x_j,y_j} \leq 1, \text{ for every eNodeB } j \in \{1, \dots, M\}, \text{ and } \text{Position}_{j,x_j,y_j} \in \{0,1\} \quad (8)$$

$$\text{Min}(\text{Cov}^{DL}(\text{Sol}), \text{Cov}^{UL}(\text{Sol})) \geq \text{Thresh1} \quad (9)$$

$\text{Position}_{j,x_j,y_j}$: This variable is equal to 1 if the eNodeB with ID *j* is installed at the location (x_j, y_j); otherwise, it is equal to 0.

$\text{Cost}_{eNB}(x_j, y_j, \text{config}(j))$ indicates the installation cost of an eNodeB; it depends on the location (x_j, y_j) and the configuration $\text{config}(j)$. For the sake of simplicity, this cost coefficient is set to 1 in our experiments, which means that all eNodeBs have the same cost; $\text{config}(j)$ involves the parameters of height, tilt, and transmit power.

The constraint defined in (8) imposes the fact that an eNodeB is at most assigned to a single location.

The constraint defined in (9) means that the coverage rate in the uplink (UL) and downlink (DL) must exceed a given threshold *Thresh1* (which can be interpreted as the target coverage rate).

$$Cov^{DL}(Sol) = \frac{1}{N} \sum_{all (x,y) \in Area} Pr_{coverage}^{DL}(x,y) \quad (10)$$

$$Cov^{UL}(Sol) = \frac{1}{N} \sum_{all (x,y) \in Area} Pr_{coverage}^{UL}(x,y) \quad (11)$$

In Equation (10), we define the DL coverage rate as the average probability of getting a satisfying SINR in the DL direction over all locations (x, y) . Likewise, Equation (11) defines the UL coverage rate as the average probability of getting a satisfying SINR in the UL direction over all possible locations (x, y) .

For the DL direction, the coverage probability at a given location (x, y) is defined as follows:

$$Pr_{coverage}^{DL}(x,y) = \begin{cases} 1, & \text{if } SINR(x,y) \geq Thresh2 \\ 0, & \text{elsewhere} \end{cases} \quad (12)$$

According to Equation (12), the coverage probability (in the DL direction) is ensured at the location (x, y) if the SINR level exceeds another threshold, *Thresh2* (which is defined in Section 5).

The SINR, as defined in Equation (13), includes all the components of large-scale path loss and the other gains in the numerator part. The denominator defines the interference power, which is caused by the neighboring BS working on the same frequency band, and it is defined in Equation (14), as well as the noise power (see Equation (15)).

The SINR in the DL direction at location (x, y) , which belongs to the cell covered by BS j over the subcarrier i , is defined in the linear scale as follows:

$$SINR(x,y) = \frac{P_{i,j}^{DL} G_t G_r LSF_{(x,y),i,j}}{I_{i,x,y} + Noise} \quad (13)$$

The numerator is the received power at the location (x, y) from the j -th BS. More specifically, $P_{i,j}^{DL}$ is the transmitted power by the BS j on the subcarrier i ; G_t, G_r are the antenna gains of the transmitter and receiver (respectively). $I(i, x, y)$ is the interference power (over subcarrier i) coming from the neighboring BS of j , and Noise is the noise power over subcarrier i (in case of UL direction, Noise is computed over another UL bandwidth).

$$I_{i,x,y} = \sum_{v \neq j} P_{i,v}^{DL} G_t G_r LSF_{(x,y),i,v} \quad (14)$$

$I(i, x, y)$ is the sum of powers transmitted by the neighboring BS of the current j -th BS over the same sub-bandwidth i .

$$Noise = KTB \quad (15)$$

where K is the Boltzmann's constant ($1.38 \times 10^{-23} \text{ J/K}$), T is the temperature in Kelvin, and B is the bandwidth in Hz (which corresponds to the sub-carrier i in DL).

In the LTE context, SINR is not computed using the power of a single subcarrier, but it is computed using RSRP, and RSRP is the average power of all reference signals in all subcarriers of all resource blocks ($12 \times \text{NRB}$), where NRB is the number of resource blocks; in our experimental study (Section 5), NRB is set to 50. In an ideal case scenario (almost zero interference and noise), and under full load, RSRP (in linear scale) is the Received Signal Strength Indicator (RSSI) divided on ($12 \times \text{NRB}$). Where, RSSI represents the total received power in the entire bandwidth ($12 \times \text{NRB}$ subcarriers).

The DL SINR in dB scale is defined as

$$SINR_{DB}(x,y) = 10 \log_{10} (P_{i,j}^{DL}) + G_t + G_r + LSF_{(x,y),i,j} - 10 \log_{10} (I + Noise) \quad (16)$$

Similarly, we compute $Pr_{coverage}^{UL}(x, y)$ using the same equation defined in (12), except that the SINR in the UL direction is computed as follows

$$SINR_{UL(DB)}(x,y) = 10 \log_{10} (P_{UE}^{UL}) + G_t + G_r + LSF_{(x,y_j),UL,(x,y)} - 10 \log_{10} (I_{UE' \neq UE} + Noise) \quad (17)$$

We mainly change the target frequency channel to UL spectrum instead of the subcarrier i , the transmit power is set to that of UE (i.e., Pr_{UE}^{UL}), the transmitter height is set to that of UE, the receiver height is set to that of BS j (this permutation is involved in the computation of LSF), the interference power ($I_{UE' \neq UE}$) is related to the other UE' that use the same UL bandwidth, and the UL noise (Noise) is computed using Equation (15) and the UL bandwidth.

4. PROPOSED APPROACH

GWO [15] is an innovative metaheuristic algorithm that simulates the hunting behavior of a pack of grey wolves. A swarm of grey wolves is structured into a social hierarchy that includes α (the leader or the best solution), β (the assistant of α or the second-best solution), δ (the helper of α and β , or the third-best solution), and ω (which represents the rest of the wolves or the remaining solutions). The main idea of the hunting process consists of three steps that are repeated throughout the iterations: encircling, search (exploration), and attack (exploitation). The algorithm leverages a perturbed distance that allows the gradual approach toward the best solution; moreover, GWO uses a hyperparameter denoted as A so as to control the trade-off between the exploration (searching for new regions) and exploitation (focusing the search on a specific region). Since all the used input variables are discretized, we adopt a discretized version of GWO (termed DGWO) that copes with our setting. The DGWO implementation assumes the following input variables:

- BSLOC: A vector containing the physical locations (x, y) of M eNodeBs. (x, y) represents the latitude and longitude of one eNodeB.
- BSH: A vector containing the heights of M eNodeBs (see Table 3 for the possible values). Each eNodeB is

characterized by three values, since we have three cells.

- **BST**: A vector containing the tilts of M eNodeBs (see Table 3 for the possible values). Each eNodeB is characterized by three values, since we have three cells.
- **NL, NC**: The number of lines and columns of the discretized study area (for coverage computation, each row or column of the grid represents around 50 m in the physical area). From these parameters, we infer the quantity N that stands for the total number of possible UE locations in the studied area: $N=NL \times NC$ (see also Equation (6)).
- **MaxM**: The maximum number of eNodesBs (or BSs), while M is the actual number of BSs.
- **BSTP**: A vector containing the transmit power of the M eNodesBs (the three sector BS).
- **COVT**: The percentage that represents the coverage target in the studied area (e.g., 95%).
- **SINRTH**: The minimum accepted SINR threshold (in dB) in every location (x, y) of the studied area.
- **P**: The size of the population (Pop) of grey wolves.

We assume that transmit power is the same for all eNodeBs (all BSTP(i) are equal); moreover, we assume that the azimuth values are the same for all eNodeBs. The three sector antennas have the following azimuth values: 0, 120, and 240. Accordingly, the optimization algorithm will tune the parameters of BSLOC, BSH, and BST for each eNodeB to get the best performance. In addition, DGWO uses a local function (see line 4) that computes the (E-UTRA Absolute Radio Frequency Channel Numbers (EARFCNs) of the DL/UL carrier frequencies and applies a (fractional) frequency reuse scheme (FFR) to plan the frequency allocation of cells. For the DGWO outputs, we assume the following quantities:

- **Toll**: The deployment cost of the best wolf (α).
- **COV**: The coverage percentage achieved by the best wolf (α).
- **W***: The configuration/position of eNodeBs.

The pseudocode of DGWO is given below:

DGWO

Input: **Pop of wolves** = $\{W_1, \dots, W_p\}$
MaxM, NL, NC, COVT, SINRTH, P

Output: **COST, COV**

1. $M=28$; $COV=0$;
2. **While** $M \leq \text{MaxM}$ **and** $COV < COVT$
3. $Pop = \text{RandomInitialization}(P, M)$
4. $\text{FrequencyAllocation}(Pop)$
5. $\alpha, \beta, \delta = \text{CoverageRanking}(Pop)$; $W^* = \alpha$
6. $A = \text{init}()$; $C' = \text{init}()$

7. **For** $t=1, Tmax$
 8. **For** $i=1, P$
 9. $D_{i1} = \text{PDist}(C', W_i, \alpha)$;
 $D_{i2} = \text{PDist}(C', W_i, \beta)$;
 $D_{i3} = \text{PDist}(C', W_i, \delta)$
 10. $X_1 = \text{Move}(A, D_{i1}, \alpha)$;
 $X_2 = \text{Move}(A, D_{i2}, \beta)$;
 $X_3 = \text{Move}(A, D_{i3}, \delta)$
 11. $W_i = X_1$ (with probability p_1)
 12. $W_i = X_2$ (with probability p_2);
 $W_i = X_3$ (with probability p_3);
 13. **End**
 14. $A = \text{update}(A)$; $C' = \text{update}(C')$
 15. $\alpha, \beta, \delta = \text{CoverageRanking}(Pop)$;
 $W^* = \text{update}(W^*, \alpha)$
 16. $CovDL = Cov^{DL}(W^*)$
 17. $CovUL = Cov^{UL}(W^*)$
 18. $COV = \text{Min}(CovDL, CovUL)$
 19. $TOLL = \text{TotalCost}(W^*)$
 20. **End**
 21. $M = M + 1$
 22. **End**
 23. **Return** $(W^*, COV, TOLL)$
-

Line 1: Initialize the values of M and COV ; the possible values of M are shown in Table 2. $MaxM$ is the highest possible value of M .

Lines 2–22: The while loop first applies the cell frequency planning (using a frequency reuse scheme) to the $M \times 3$ cells and then invokes the discrete GWO to ensure the target coverage rate $COVT$; if $DGWO$ fails (in achieving $COVT$), M is incremented (see line 21), and we try another $DGWO$ simulation.

Line 3: We randomly initialize the P wolves; each wolf is a quadruplet $W_i = (BSLOC, VH, VT, BSTP)$; it represents a positioning/configuration of all $eNodeBs$. This step initializes the M $eNodeBs$ using the domains shown in Table 3.

Line 4: We assign the frequency channels to the cells (using $EARFCNs$ and the frequency reuse scheme).

Line 5: We sort the wolves of Pop (using the coverage metric) and retain the Top 3 solutions. This metric is computed in Line 18, and it is based on Equations (10) and (11) cited in Section 3. In addition, the best solution W^* is initialized (i.e., $W^* = \alpha$).

Line 6: In our discrete version, C' is a binary vector that indicates the wolf's dimensions (C' and W_i have the same size) that are involved in the distance computation shown in Line 9. Moreover, the trade-off factor A is a vector of the same size as C' ; each element $A(i)$ of A belongs to $[-1, 1]$ and it is linearly decremented toward 0 (see line 14).

Line 7: It is the principal loop of *GWO* that controls the wolves' position updates.

Lines 8–13: This loop iterates over all wolves and performs several tasks: (1) the computation of perturbed distance D_{ij} (see Line 9); (2) the computation of the temporary position of W_i according to A , D_{ij} , and the Top_j wolf of the population (see Line 10); and (3) the effective value of W_i that is randomly selected from the three candidate positions calculated in step (2) (see Lines 11–12).

The selection probabilities P_1 , P_2 , and P_3 are tuned during the experiments (see Section 5).

Line 14: The hyperparameter C is randomly updated; the hyperparameter A is linearly decreased toward 0.

Line 15: The Top3 wolves, as well as W^* , are updated.

Lines 16–18: The coverage rate of W^* is computed using the *DL* and *UL* orientations.

Line 19: The total cost of W^* is computed.

Line 21: M is updated.

Line 23: The best solution, the total cost, and the coverage rate are returned to the decision-maker.

The improved GWO (termed IGWO) shown in the next pseudocode consists of replacing the update based on the δ wolf by a local search operator. More specifically, in Line 11 of IGWO, we first choose a random wolf from the set $\{\alpha, \beta, \delta\}$, and then we apply a local search on that wolf by choosing two dimensions to update.

If the chosen dimension is the height of a given eNodeB, then the old value $V_k \in \{V_1, V_2, \dots, V_k\}$ is replaced by V_{k-1} or V_{k+1} .

The same thing can be said for the tilt dimensions. The old value is replaced by one of the two neighboring values in the variable domain. If the chosen dimension is the physical location (x, y) of a given eNodeB, then

$$(x_{new}, y_{new}) = (x_{old}, y_{old}) + (\Delta_x, \Delta_y) \quad (18)$$

where

$\Delta_x \in \{-50m, -25m, 0m, 25m, 50m\}$, and
 $\Delta_y \in \{-50m, -25m, 0m, 25m, 50m\}$.

The adopted resolution step for placing eNodeB is set to 25 m.

The remaining instructions of IGWO are the same as for DGWO. In Section, we evaluate the impact of this new operator on both the cost and coverage rate.

It is worth noting that the SINR formula (see Equations (13) and (16)) is learned using the Adaboost method [51]. In particular, we generated hundreds of pairs of inputs/outputs using the Atoll simulator (version 3.3.2). The input of each pair (or example) contains the 03 vectors BSLOC, BSH, and BST (the eNodeB positioning and configuration), while the output contains an $NL \times NC$ matrix of SINR values generated by the input disposition of eNodeBs. We must mention that Adaboost is based on the KNN regression weak -learner.

IGWO

Input: **Pop of wolves** = $\{W_1, \dots, W_P\}$
MaxM, NL, NC, COVT, SINRTH, P

Output: **COST, COV**

1. $M=28; COV=0;$
 2. **While** $M \leq MaxM$ **and** $COV < COVT$
 3. $Pop = RandomInitialization(P, M)$
 4. $FrequencyAllocation(Pop)$
 5. $\alpha, \beta, \delta = CoverageRanking(Pop); W^* = \alpha$
 6. $A = init(); C' = init()$
 7. **For** $t=1, Tmax$
 8. **For** $i=1, P$
 9. $D_{i1} = PDist(C', W_i, \alpha);$
 $D_{i2} = PDist(C', W_i, \beta);$
 10. $X_1 = Move(A, D_{i1}, \alpha);$
 $X_2 = Move(A, D_{i2}, \beta);$
 11. $Wolf = RandomChoice(\alpha, \beta, \delta);$
 $X_3 = LocalSearch(Wolf)$
 12. $W_i = X_1$ (with probability p_1);
 13. $W_i = X_2$ (with probability p_2);
 14. **End**
 15. $A = update(A); C' = update(C')$
 16. $\alpha, \beta, \delta = CovergeRanking(Pop);$
 $W^* = update(W^*, \alpha)$
 17. $CovDL = Cov^{DL}(W^*)$
 18. $CovUL = Cov^{UL}(W^*)$
 19. $COV = Min(CovDL, CovUL)$
 20. $TOLL = TotalCost(W^*)$
 21. **End**
 22. $M = M + 1$
 23. **End**
 23. **Return**($W^*, COV, TOLL$)
-

To evaluate the time complexity of IGWO, we first give insights about the complexity of the standard GWO, frequency planning, SINR complexity, and local search operator.

We notice that the standard GWO has a complexity of $O(Tmax * P * D_{in})$, where:

D_{in} (dimensionality of the input): it is estimated as $NumberOfCells * 4$, (4 represents the number of parameters, which are the location (x, y) of BS, height, and tilt).

NumberOfCells: it is the number of eNodeB times 3 ($M * 3$).

MaxM: it the maximum number of eNodeBs, in our experiments it is set to 42.

Tmax: the number of iterations (see the possible values in Table 2).

P : the size of wolf population (see the possible values in Table 2).

D_{out} (dimensionality of the output): it is a 2D matrix of $N=11464$ floats, each element represents the SINR of a square region of $50*50$ m².

$NumberOfChannels$ (NC): we divided the spectrum of the DL stream of E-UTRA Band3 into $NC=6$ channels, each channel is assigned to a cell in a way that minimizes the interference.

We also notice that the step 4 of IGWO (frequency planning) has a complexity of $O(NC \cdot NumberOfCells)$.

The SINR calculation done in lines (17 or 18) is achieved using the adaboost testing phase which has a cost equal to $O(NumberOfweaklearner * TrainingSetSize * D_{in} * D_{out})$.

$NumberOfWeakLearners$: by default, it is set to 3.

$TrainingSetSize$: by default, it is set to 500 examples (it represents the cardinal of the training set).

The complexity of the local search (line 11 of IGWO) is the product of the SINR complexity with neighborhood size.

We define the neighborhood size (NS) of a solution as $NS=SizeTilts * SizeHeights * SizePositions$. To reduce the computation overhead, we consider only one new position (for $SizePositions$) instead of 5 possible positions. Thus, $NS=8*10*1=80$. The local search complexity is:

$$O(NS * NumberOfweaklearner * TrainingSetSize * D_{in} * D_{out}).$$

The complexity of IGWO (with frequency planning) is:

$$O(MaxM * (P * MaxM + NC^{NumberOfCells} + (P * NumberOfweaklearner * TrainingSetSize * D_{in} * D_{out}) + T_{max} * (P * (NumberOfCells + (NS * NumberOfweaklearner * TrainingSetSize * D_{in} * D_{out}))) + (NumberOfweaklearner * TrainingSetSize * D_{in} * D_{out}))).$$

5. RESULTS AND DISCUSSIONS

The reported experiments were conducted using a machine with an Intel Core i5-1245U CPU at 1.60 GHz, 16GB memory (RAM), and Windows 12 with a 64-bit operating system. We used Python 3.9 to develop all the algorithms, namely DGWO, DGWO with a single leading agent (AlphaGWO), IGWO, DPSO, and TS [52]. In AlphaGWO, we only use the Alpha agent, the beta and delta agents are eliminated, this variant serves for evaluating the impact of the voting rule on the performance of the search.

We simulated the planning on an urban geographic area (Oran city of Algeria); more specifically, we chose a rectangular area with sides measuring 4 km and 7 km for an area of 28 km². To compute the SINRs (Equations (13) and (16)), we used the Cost-231-Hata empirical model. Initially, the eNodeBs were randomly distrib-

uted in the area according to the uniform law with a density of 1 eNodeB per km²; thus, we started the simulation with $M = 28$ eNodeBs (and consequently we started with $28*3 = 84$ cells). This number was gradually increased so as to meet the target coverage rate.

Our experiments adopt the E-UTRA Band 3 - 10MHz band. We defined 6 channels for allocating the spectrum to eNodeBs (in particular, we have 6 DL-EARFCNs).

In the following, we give the frequency plan for DL-EARFCNs. For the DL orientation (1805-1880 MHz), we defined the first and the last EARFCNs as follows:

First DL EARFCN= 1 250, last DL EARFCN = 1 850, the step separating the channels is set to 100.

Table 2 shows the technical parameters of our two algorithms (DGWO and IGWO). Table 3 shows the simulation parameters that concern the deployed eNodeB (e.g., frequency band, channel bandwidth, and antenna gain). Furthermore, we assume that the UE antenna gain is set to 1 in the linear scale (or 0 dBi).

We also suppose that the resolution step in Equation (6) is equal to 50 m for calculating the coverage formula (see Equations (10) and (11)); this variable determines the SINR matrix size; however, it is set to 25 m for placing the eNodeBs (this setting determines the minimum values of $(\Delta x, \Delta y)$).

To initialize the wolves' positions/configurations of DGWO and IGWO, we first generated a hundred random values; thereafter, we selected the first P wolves that had the best coverage score; these P wolves constitute the first generation of the wolves' population.

Table 2. GWO Parameters

Parameter	Values
Population size (P)	20, 30, 40, 50
Max number of iterations (T_{max})	20, 30
Number of eNodeBs (M)	28, 32, 36, 38, 40, 42
Probability thresholds P_1, P_2, P_3	40%, 30%, 30%
Target coverage rate ($COVT$)	95%
SINR threshold	10 dB, 12 dB

Table 3. Simulation parameters.

Parameter	Values
Antenna transmit power	43 dBm
Shadowing standard deviation	5 dB
UL and DL channel bandwidth	10 MHz
Frequency band	1800 MHz
MIMO configuration	2,2 MIMO
eNodeB Antenna gain	17 dBi
eNodeB Tilt	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
eNodeB Height	10, 15, 20, 25, 30, 35, 40, 45
UE Height	1.5

Fig. 1 shows a comparison between IGWO and GWO in terms of coverage (with a minimum SINR of 10 dB), $M = 36$, and $P = 30$. The results confirm the superiority of IGWO with respect to DGWO. We observe that both of

them converge around the fifth or sixth iteration. However, we note that IGWO can reach 98.8% of 10 dB SINR satisfaction, while DGWO reaches only 88.2% of satisfaction rate (Thresh2 = 10 dB). Therefore, we conclude from this experiment that the addition of a local search (applied on the Top3 wolves) has a positive impact on the coverage performance. This operator modifies a subset of parameters using the neighborhood values and aims at enhancing the wolf score. In the experiments introduced in Figs. 2–6, we assume that $M = 36$.

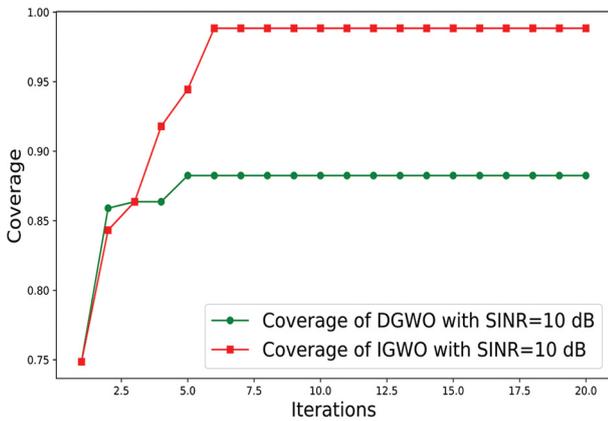


Fig. 1. Coverage rates for both DGWO and IGWO (Thresh2 = 10 dB)

Fig. 2 demonstrates the coverage evolution of IGWO (of the previous execution) for two different SINR thresholds. We first observe that the convergence is ensured after six iterations (with a population of 30 wolves), then we notice that the ensured coverage rate is 98.8% for a threshold of 10 dB; however, it is equal to 76.0% for a SINR threshold of 12 dB. This experiment indicates the necessity of adding more eNodeBs to reach the target coverage for 12 dB SINR.

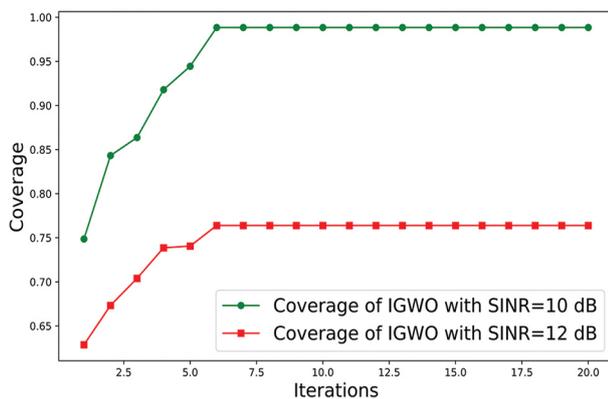


Fig. 2. Coverage rates of IGWO for Thresh2 = 10 dB and Thresh2 = 12 dB

Fig. 3 shows the impact of the population size on the average coverage percentages. It is clearly indicated that a size P greater than or equal to 30 can achieve satisfying results for a threshold of 10 dB, while the average coverage rate is still low for 12 dB (for all P) due to the weak value of M .

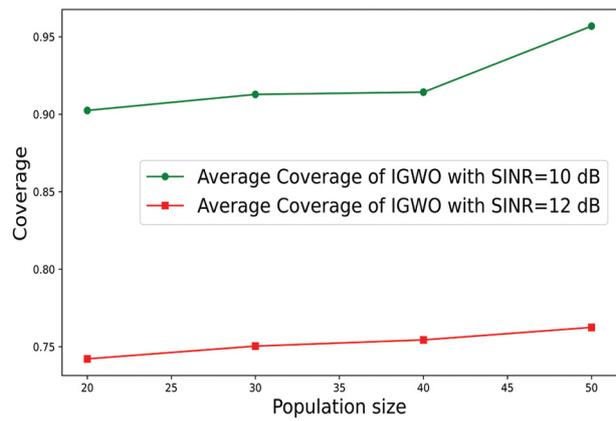


Fig. 3. Average coverage rate vs. Population size

Fig. 4 illustrates the consumed CPU time with respect to the population size of IGWO and the size of the neighborhood of TS; the number of iterations is set to 20 for both methods. We notice a linear increase in time for both techniques; in particular, the CPU time ranges from 400 s up to 1100 s for IGWO, while it ranges from 800 s to 1800 s for TS. Accordingly, we conclude that the perfect population size (of IGWO) ensuring an acceptable delay for the user ranges between 30 and 40.

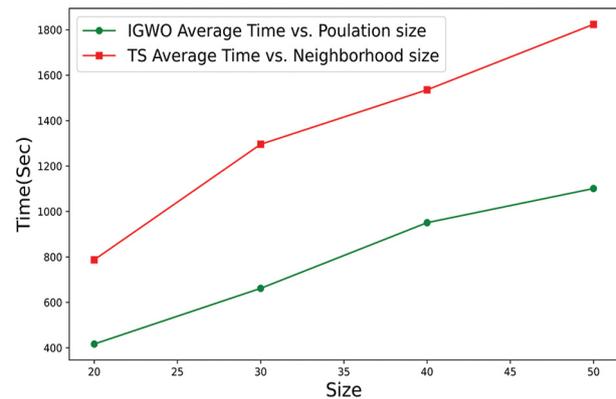


Fig. 4. IGWO execution time vs. Population size

In Fig. 5, we show a comparison between the performance of IGWO, DPSO, AlphaGWO, and TS. We notice that IGWO, DPSO, and AlphaGWO have the same population size which set to 30 ($P = 30$). Clearly, IGWO outperforms all methods for the same SINR threshold (10 dB). We note that TS was implemented with a neighborhood size of 40 and a Tabu list size of 3. Despite the fact that both algorithms (TS and IGWO) converge before the sixth iteration, we notice that the local search used in TS is not sufficient to optimize the coverage of a moderately sized geographical area (despite using a large neighborhood). The experiment shows that the TS simulation reaches a coverage rate of 88.3% for a SINR threshold of 10 dB. On the other hand, we notice that DPSO is ranked second, and consequently this social-oriented gradient descent confirms its effectiveness in cellular planning (it reaches a coverage rate of 96.2 %). Regarding AlphaGWO, we observe that its coverage rate is slightly lower than that of TS (and even DGWO).

In fact, AlphaGWO reaches a coverage rate of 88.1%, and therefore, we can conclude that the voting rule (which is implemented in DGWO and absent in AlphaGWO) has a powerful impact on the quality of the derived solutions.

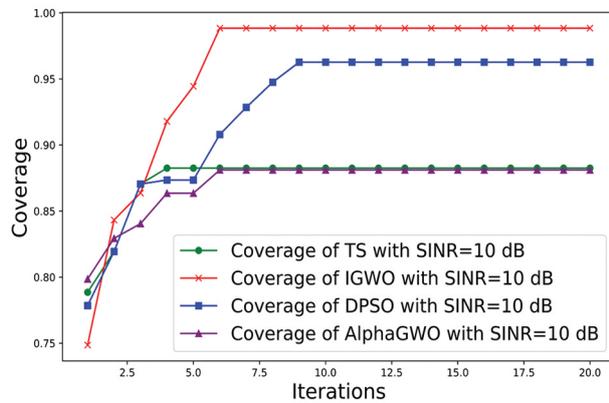


Fig. 5. IGWO performance vs. All methods

In Fig. 6, we present the frontiers of cells ($36 \times 3 = 108$) and their respective SINR levels in the studied urban area. This result corresponds to the best wolf given by the IGWO experiments illustrated in Figures 1 and 2 (i.e., 98.8% of 10 dB SINR satisfaction or 76.0% of 12 dB SINR satisfaction). The green color represents a SINR of at least 10 dB, and the yellow and red colors represent higher SINRs.

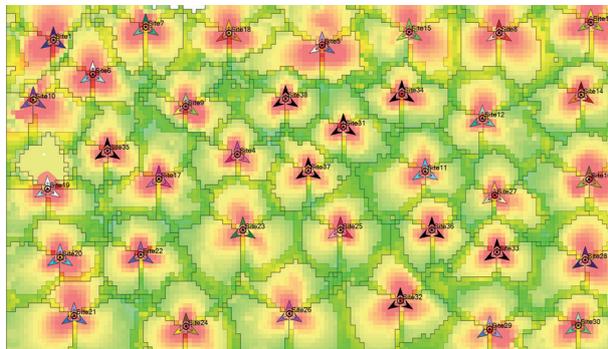


Fig. 6. Cell frontiers for the best wolf given by IGWO with $M = 36$

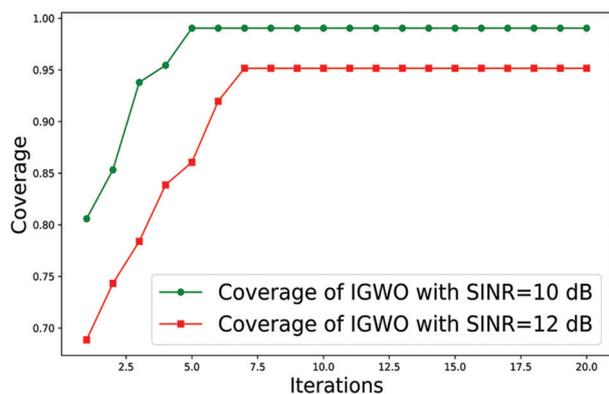


Fig. 7. Coverage rates of IGWO with $M=42$, Thresh2=10 db, and Thresh2=12 db

In Fig. 7, we present the evolution of IGWO with $M = 42$ and $P = 30$. As clearly indicated, IGWO converges to the final performance of 95.1% after seven iterations (using a SINR threshold of 12 dB). However, it reaches the performance of 99.0% for the 10 dB SINR threshold after five iterations. This experiment supports the efficiency of IGWO in reaching near-optimal solutions of eNodeB deployment.

6. CONCLUSION

We have presented in this paper an automatic approach for planning the coverage of an LTE network while considering compromises in cost and capacity. Our proposition includes the development of both a discrete GWO and an improved GWO for tuning the parameters of eNodeBs. Besides replacing the averaging rule with the probabilistic voting rule, the improved version of the Grey Wolves Optimizer (IGWO) contains an additional local search operator during the exploration phase. This operator modifies a subset of parameters using the neighborhood values and is aimed at enhancing the wolf score. Our algorithm optimizes the eNodeBs positions, their heights, and their tilts and provides a deployment option with the best coverage/cost pair. We also highlight that IGWO is able to outperform the standard GWO, the PSO technique, and the TS metaheuristic in coverage planning based on the SINR threshold of 10 dB.

In future works, we plan to compare our proposition with other effective metaheuristics such as Dragonfly algorithm or spider monkey optimization (SMO). Additionally, we aim to handle other objective functions (e.g., number of users and power consumption) or other constraints. Finally, we also aim to extend our work for planning other emergent networks.

7. REFERENCES

- [1] A. Qodirov, U. Amirsaidov, "Cross-Layer Model of Dynamic Distribution of Radio Resources and Data Flow Service in LTE Networks", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 1, 2023, pp. 13-19.
- [2] M. J. Alam, M. R. Hossain, S. Azad, R. Chugh. "An overview of LTE/LTE-A heterogeneous networks for 5G and beyond", Transactions on Emerging Telecommunications Technologies, Vol 34, No. 8, 2023, p. 4806.
- [3] E. Dahlman, S. Parkvall, J. Skold, "4G: LTE/LTE-advanced for mobile broadband", Academic Press, 2013.
- [4] J. Premalatha, A. S. A. Nisha, S. E. S. Pillai, A. Bhuvanesh, "Performance Measurement of Small Cell Power Management Mechanism in 5G Cellular

Networks using Firefly Algorithm”, *International Journal of Electrical and Computer Engineering Systems*, Vol. 15, No. 5, 2024, pp.437-48.

- [5] J. Gandhi, Z. Narmawala, “Coverage, capacity and cost analysis of 4g-lte and 5g networks: A case study of Ahmedabad and Gandhinagar”, *Proceedings of the Futuristic Trends in Networks and Computing Technologies: Select Proceedings of Fourth International*, November 2022, pp. 25-38.
- [6] D. Tsilimantos, D. Kaklamani, G. Tsoulos, “Particle swarm optimization for UMTS WCDMA network planning”, *Proceedings of the 3rd International Symposium on Wireless Pervasive Computing*, Santorini, Greece, 7-9 May 2008, pp. 283-287.
- [7] I. K. Valavanis, G. Athanasiadou, D. Zarbouti, G. V. Tsoulos, “Base-station location optimization for LTE systems with genetic algorithms”, *Proceedings of European Wireless; 20th European Wireless Conference*, Barcelona, Spain, 14-16 May 2014, pp. 1-6.
- [8] E. Yaacoub, Z. Dawy, “LTE radio network planning with HetNets: BS placement optimization using simulated annealing”, *Proceedings of the MELECON 2014 - 2014 17th IEEE Mediterranean Electrotechnical Conference*, Beirut, Lebanon, 13-16 April 2014, pp. 327-333.
- [9] E. Yaacoub, A. Imran, Z. Dawy, “A generic simulation-based dimensioning approach for planning heterogeneous LTE cellular networks”, *Proceedings of the MELECON 2014 - 2014 17th IEEE Mediterranean Electrotechnical Conference*, Beirut, Lebanon, 13-16 April 2014, pp. 320-326.
- [10] M. S. Sharawi, “RF planning and optimization for LTE Networks”, *Evolved Cellular Network Planning and Optimization for UMTS and LTE*, CRC Press, 2010, pp. 399-432.
- [11] M. Galota, C. Glaer, S. Reith, H. Vollmer, “A polynomialtime approximation scheme for base station positioning in UMTS networks”, *Proceedings of the 5th international workshop on Discrete algorithms and methods for mobile computing and communications*, Rome, Italy, July 2001, pp. 52-59.
- [12] J. Isabona, A. L. Imoize, S. Ojo, P. Venkatareddy, S. K. Hinga, M. Sánchez-Chero, S. M. Ancca, “Accurate base station placement in 4G LTE networks using multiobjective genetic algorithm optimization”, *Wireless Communications and Mobile Computing*, Vol. 2023, No. 1, 2023, pp. 1-9.
- [13] I. Vatsh, V. Gupta, B. Bhattacharyya, “Optimizing base station deployment for LTE using metaheuristic algorithms”, *Proceedings of the IEEE 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking*, Tamilnadu, India, 30-31 March 2019, pp. 1-5.
- [14] F. Luna, R. M. Luque-Baena, J. Martinez, J. F. Valenzuela-Valdés, P. Padilla, “Addressing the 5G cell switch-off problem with a multi-objective cellular genetic algorithm”, *Proceedings of the IEEE 5G World Forum*, Silicon Valley, CA, USA, 9-11 July 2018, pp. 422-426.
- [15] S. Mirjalili, S. M. Mirjalili, A. Lewis. “Grey wolf optimizer”, *Advances in Engineering Software*, Vol. 69, 2014, pp. 46-61.
- [16] J. Brennan, “Condorcet's Jury Theorem and the Optimum Number of Voters”, *Politics*, Vol. 31, No. 2, 2011, pp. 55-62.
- [17] R. Borralho, A. Mohamed, A. U. Quddus, P. Vieira, R. Tafazolli, “A survey on coverage enhancement in cellular networks: Challenges and solutions for future deployments”, *IEEE Communications Surveys & Tutorials*, Vol. 23, No. 2, 2021, pp. 1302-1341.
- [18] A. Taufique, M. Jaber, A. Imran, Z. Dawy, E. Yacoub, “Planning wireless cellular networks of future: Outlook, challenges and opportunities”, *IEEE Access*, Vol. 5, 2017, pp. 4821-4845.
- [19] R. S. Hassan, T. A. Rahman, A. Y. Abdulrahman, “LTE coverage network planning and comparison with different propagation models”, *Telkomnika (Telecommunication Computing Electronics and Control)*, Vol. 12, No. 1, 2014, pp. 153-162.
- [20] O. O. Erunkulu, A. M. Zungeru, C. K. Lebekwe, J. M. Chuma, “Cellular communications coverage prediction techniques: A survey and comparison”, *IEEE Access*, Vol. 8, 2020, pp. 113052-113077.
- [20] H. Elsayy, A. Sultan-Salem, M. Alouini, M. Z. Win, “Modeling and analysis of cellular networks using stochastic geometry: A tutorial”, *IEEE Communications Surveys & Tutorials*, Vol. 19, No. 1, 2017, pp. 167-203.

- [21] H. ElSawy, E. Hossain, M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey", *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 3, 2013, pp. 996–1019.
- [22] A. Imran, E. Yaacoub, Z. Dawy, A. Abu-Dayya, "A mathematical modeling approach and a novel solution for sector azimuth angle planning", *Proceeding of the MELECON 2014 - 17th IEEE Mediterranean Electrotechnical Conference*, Beirut, Lebanon, 13-16 April 2014, pp. 334-338.
- [23] S. K. Jha, R. Rokaya, A. Bhagat, A. R. Khan, L. Aryal, "LTE network: Coverage and capacity planning—4G cellular network planning around Banepa", *Proceeding of the IEEE International Conference on Networking and Network Applications*, Kathmandu City, Nepal, 16-19 October 2017, pp. 180-185.
- [24] R. Nlend, E. Tonye. "Planning and simulation of LTE radio network: case of the city of Yaoundé", *IOSR Journal of Electronics and Communication Engineering*, Vol. 14, No. 2, 2019, pp. 19-29.
- [25] A. Idris, S. Rahmatia, M. Ismail, "4G LTE Cellular Network Coverage Planning and Simulation on Mandalay Area with Propagation Model Cost-Hata", *Proceeding of IEEE 9th International Conference on Information and Communication Technology*, Yogyakarta, Indonesia, 3-5 August 2021, pp. 280-285.
- [26] V. F. Guijarro, J. D. Vega-Sánchez, M. P. Paredes, F. G. Arévalo, D. M. Osorio, "Comparative Evaluation of Radio Network Planning for Different 5G-NR Channel Models on Urban Macro Environments in Quito City", *IEEE Access*, Vol. 12, 2024, pp. 5708-5730.
- [27] A. Yuhaneef, S. Aulia, J. Jasmanto, "Optimization of 4G LTE Network Bad Spot Area using Automatic Cell Planning Method", *Jurnal ELTIKOM: Jurnal Teknik Elektro, Teknologi Informasi dan Komputer*, Vol. 7, No. 2, 2023, pp. 181-191.
- [28] A. Imran, M. Imran, R. Tafazolli, "Relay station access link spectral efficiency optimization through SO of Macro BS tilts", *IEEE Communications Letters*, Vol. 15, 2011, pp. 1326-1328.
- [29] Y. Qi, M. Imran, R. Tafazolli, "Energy-aware adaptive sectorization in LTE systems", *Proceeding of the IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications*, Toronto, ON, Canada, 11-14 September 2011, pp. 2402-2406.
- [30] A. Imran, M. A. Imran, R. Tafazolli, "A novel self organizing framework for adaptive frequency reuse and deployment in future cellular networks", *Proceeding of the IEEE 21st International Personal Indoor and Mobile Radio Communications Symposium*, Istanbul, Turkey, 26-30 September 2010, pp. 2354-2359.
- [31] M. Aldajani, "Convolution-based placement of wireless base stations in urban environment", *IEEE Transactions on Vehicular Technology*, Vol. 57, No. 6, 2008, pp. 3843-3848.
- [32] R. Nlend, E. Tonye, "Planning and optimization approach using genetic algorithms of a new generation Cellular network capitalizing on the existing sites", *International Journal of Science and Research*, Vol. 8, No. 5, 2019, pp. 1419-1427.
- [33] E. Nwelih, J. Isabona, A. L. Imoize, "Optimization of Base Station Placement in 4G LTE Broadband Networks Using Adaptive Variable Length Genetic Algorithm", *SN Computer Science*, Vol. 4, No. 2, 2022, p. 121.
- [34] V. Berrocal-Plaza, M. Vega-Rodriguez, J. Gomez-Pulido, J. Sanchez-Perez, "Artificial bee colony algorithm applied to WiMAX network planning problem", *Proceeding of the 11th International Conference on Intelligent Systems Design and Applications*, Córdoba, Spain, 22-24 November 2011, pp. 504-509.
- [35] S. Debnath, W. Arif, D. Sen, S. Baishya, "LTE Cell Planning for Resource Allocation in Emergency Communication", *Wireless Personal Communications*, Vol. 135, No. 2, 2024, pp. 1035- 1076.
- [36] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems", *Neural Computing and Applications*, Vol. 27, No. 4, 2016, pp. 1053-1073.
- [37] M. Komala, I. Wahidah. "LTE networks BTS location optimization with double step grey Wolf optimizer", *Proceeding of the IEEE 11th International Conference on Telecommunication Systems Services and Applications*, Lombok, Indonesia, 26-27 October 2017, pp. 1-6.

- [38] A. Al-Samawi, A. Sali, N. K. Noordin, M. Othman, F. Hashim, "Base station location optimization in LTE using genetic algorithm", *Proceeding of the IEEE International Conference on ICT Convergence*, Jeju Island, Korea, 14-16 October 2013, pp. 336-341.
- [39] E. Amaldi, A. Capone, F. Malucelli. "Radio planning and coverage optimization of 3G cellular networks", *Wireless Network*, Vol. 14, No. 4, 2008, pp. 435-447.
- [40] B. Sapkota, R. Ghimire, P. Pujara, S. Ghimire, U. Shrestha, R. Ghimire, B. R. Dawadi, S. R. Joshi, "5G Network Deployment Planning Using Metaheuristic Approaches", *Telecom*, Vol. 5, No. 3, 2024, pp. 588-608.
- [41] A. Kalakech, M. Bakir, R. Youssef, "Coverage and Cell Capacity optimization in Private LTE network based on Position and Expected Channel Knowledge", *Proceedings of the International Conference on Computer and Applications*, Cairo, Egypt, 28-30 November 2023, pp. 1-6.
- [42] F. Calles-Esteban, A. A. Olmedo, C. J. Hellín, A. Valledor, J. Gómez, A. Tayebi, "Optimizing Antenna Positioning for Enhanced Wireless Coverage: A Genetic Algorithm Approach", *Sensors*, Vol. 24, No. 7, 2024, p. 2165.
- [43] R. M. Dreifuerst, S. Daulton, Y. Qian, P. Varkey, M. Balandat, S. Kasturia, A. Tomar, A. Yazdan, V. Ponnampalam, R. W. Heath, "Optimizing coverage and capacity in cellular networks using machine learning", *Proceeding of ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, ON, Canada, 6-11 June 2021, pp. 8138-8142.
- [44] Z. H. Yang, M. Chen, Y. P. Wen, L. Q. Jia, Y. Zhang, "Cell planning based on minimized power consumption for LTE networks", *Proceeding of the IEEE Wireless Communications and Networking Conference*, Doha, Qatar, 3-6 April 2016, pp. 1-6.
- [45] G. E. Athanasiadou, D. Zarbouti, G. V. Tsoulos, "Automatic location of base-stations for optimum coverage and capacity planning of LTE systems", *Proceeding of the IEEE 8th European Conference on Antennas and Propagation*, The Hague, Netherlands, 6-11 April 2014, pp. 2077-2081.
- [46] Y. Wu, S. Pierre, "A new hybrid constraint-based approach for 3G network planning", *IEEE Communications Letters*, Vol. 8, No. 5, 2004, pp. 277-279.
- [47] B. Liu, "LTE wireless network coverage optimisation based on corrected propagation model", *Applied Mathematics and Nonlinear Sciences*, Vol. 8, No.1, 2022, pp. 2681-2694.
- [48] H. Putri, I. Ahmad, A. Hikmaturokhman, D. H. Putri, "Automatic Cell Planning Method for Radio Network Optimization", *JOIV: International Journal on Informatics Visualization*, Vol. 8, No.1, 2024, pp. 6449-6455.
- [49] N. Bilal, P. Galinier, F. Guibault, "A new formulation of the set covering problem for metaheuristic approaches", *International Scholarly Research Notices*. Vol. 2013, 2013, pp. 1-10.
- [50] COST Action 231, "Digital mobile radio towards future generation systems, final report", *European Communities*, Technical Report, EUR 18957, 1999.
- [51] Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, Vol. 55, No. 1, 1997, pp. 119-39.
- [52] F. Glover, "Tabu search—part I", *ORSA Journal on Computing*, Vol. 1, No. 3, 1989, pp. 190-206.

Enhancing Energy Efficiency in GAN-based HEVC Video Compression Using Knowledge Distillation

Original Scientific Paper

Hajar Hardi*

Sciences and Techniques for the Engineer Laboratory
National School of Applied Sciences University Sultan Moulay Slimane Khouribga, Morocco
hardi.hajar@gmail.com

Imade Fahd Eddine Fatani

Sciences and Techniques for the Engineer Laboratory
National School of Applied Sciences University Sultan Moulay Slimane Khouribga, Morocco
i.fatani@usms.ma

*Corresponding author

Abstract – High-efficiency Video Coding (HEVC) is a widely used video coding standard, and it has recently gained widespread adoption in various applications, such as video streaming, broadcasting, real-time conferencing, and storage. The adoption of Generative Adversarial Networks (GANs) into HEVC compression has shown significant improvements in compression performance by reducing the video size while maintaining the original quality. In this work, we explore the application of Knowledge Distillation to reduce the energy consumption associated with GAN-based HEVC. By training a smaller student model that imitates the larger teacher model's behavior, we significantly improved energy efficiency. In this paper, we provide a detailed study comparing the traditional HEVC algorithm, GAN-based HEVC, and GAN-based HEVC with Knowledge Distillation. The experimental results demonstrate a reduction in energy consumption of up to 30% while preserving video quality, making it an effective solution for video streaming platforms and energy-constrained devices and a sustainable solution for video compression without diminishing video quality.

Keywords: HEVC, High-Efficiency Video Coding, Video Compression, GAN, Generative Adversarial Networks, Knowledge Distillation, Student-Teacher Model, Power Consumption Optimization, Energy Efficiency

Received: November 21, 2024; Received in revised form: February 11, 2025; Accepted: February 13, 2025

1. INTRODUCTION

In the past few years, video content has become omnipresent in various applications such as streaming, broadcasting, real-time conferencing, and storage. This growing demand has significantly increased the need for efficient video compression methods to optimize storage and transmission while maintaining visual quality. High-Efficiency Video Coding (HEVC) has emerged as one of the most widely adopted video compression standards due to its superior rate-distortion performance compared to its predecessors [1]. However, the pursuit of higher compression efficiency often comes at the cost of increased computational complexity.

Integrating Generative Adversarial Networks (GANs) into HEVC-based compression has introduced a paradigm shift, enhancing the perceptual quality of compressed videos by leveraging learned representations for frame re-

construction [2]. GAN-based approaches have remarkably improved subjective visual quality metrics, particularly in reducing artifacts introduced by aggressive compression [3]. However, despite their effectiveness, these methods suffer from significant computational overhead, leading to increased energy consumption, especially when deployed on resource-constrained devices such as mobile platforms and edge computing environments [4].

Several works have explored the application of deep learning models in video compression, focusing on balancing efficiency and perceptual quality [5]. However, the trade-off between computational cost and visual fidelity in GAN-based HEVC remains an open research challenge. In particular, the energy footprint of GAN-enhanced video coding has not been extensively studied, and existing methods do not adequately address the need for reducing power consumption while maintaining high-quality compression [6].

To address this limitation, we propose a novel approach that integrates knowledge distillation (KD) with GAN-based HEVC to optimize energy efficiency while preserving perceptual quality. Knowledge distillation is a widely used model compression technique where a smaller student model is trained to imitate the behaviour of a larger, computationally intensive teacher model [7]. This technique has been successfully applied in various computer vision and natural language processing tasks to reduce inference time and memory requirements [8]. However, its application in GAN-based video compression remains largely unexplored. Unlike previous works that focus solely on enhancing GAN-based compression, our method leverages KD to transfer knowledge from a high-performance GAN model to a lightweight alternative, significantly reducing energy consumption while maintaining competitive visual quality.

In this work, we conduct a comprehensive comparative analysis of traditional HEVC, GAN-based HEVC, and GAN-based HEVC with knowledge distillation evaluating energy consumption across different hardware components, including CPU (IA Energy) and GPU (GT Energy). Our experimental results demonstrate that knowledge distillation can achieve up to 30% reduction in energy consumption, offering a scalable and sustainable solution for energy-constrained video applications. By bridging the gap between energy efficiency and high-quality GAN-enhanced compression, this study introduces a practical optimization strategy for next-generation video coding systems.

2. RELATED WORKS

This section provides an in-depth analysis of recent studies related to GAN optimization, energy efficiency, and HEVC-based video compression. The focus is on understanding the advantages and limitations of each method, discussing their experimental setups, and positioning our approach within this evolving research landscape.

Recent research has explored various strategies to optimize GANs for improved computational efficiency and energy-aware processing. One notable work by [9] proposes a method for enhancing data efficiency in GANs by dynamically adjusting the network's structure during training. The key contribution of this work is its ability to reduce computational costs while maintaining image generation quality. However, the paper focuses primarily on static image generation, leaving open questions about its applicability to video compression scenarios.

Building on this, another study [10] introduces a data-efficient GAN training framework that reconfigures the GAN architecture during training. Unlike [9], which focuses on modifying network layers, this work adaptively adjusts model complexity in real-time, minimizing resource consumption while preserving perceptual quality. However, both approaches primarily target

GAN training rather than inference efficiency, which is a critical factor for deployment in real-world video streaming applications.

In the context of mobile and low-power environments, a noteworthy contribution from [11] investigates GAN energy consumption trade-offs for mobile platforms. The study proposes a lightweight GAN model that balances image generation accuracy with battery efficiency. The experimental results demonstrate up to a 25% reduction in energy usage, but the method is primarily tailored for image generation rather than video compression, limiting its direct applicability to HEVC.

Beyond optimizing GAN architectures, researchers have explored how GANs can contribute to energy efficiency in broader applications. For example, an ensemble-based approach in [12] integrates GANs to reduce energy consumption in commercial buildings. While this work does not directly focus on reducing GAN power consumption, it showcases the potential of GANs in forecasting applications, which could be leveraged for adaptive energy-aware compression strategies.

A final relevant study [13] examines the role of adaptive computation in controlling energy consumption during GAN-based generation. The key takeaway from this work is that dynamic computational allocation can significantly reduce energy costs while maintaining generation quality. However, the proposed technique primarily applies to classification and object detection tasks, requiring further adaptation for video compression pipelines.

While the above studies primarily focus on GAN energy efficiency, recent research has also explored the role of GANs in HEVC compression. One seminal study [14] investigates GAN-based enhancement for HEVC-encoded videos, demonstrating significant improvements in perceptual quality. However, the increased computational complexity limits real-time deployment.

A more recent work from 2024 [15] explores how transformer-based GANs can enhance HEVC compression by reconstructing lost details in low-bitrate videos. The study shows up to a 30% improvement in SSIM and VMAF scores, but its energy overhead is not explicitly analyzed, leaving room for energy-efficient optimizations.

To mitigate the energy cost of GANs, some studies have begun incorporating Knowledge Distillation (KD) into the compression pipeline. A recent study [16] integrates KD into GAN-based HEVC enhancement, demonstrating a 20% reduction in inference time while maintaining perceptual quality. However, this method relies heavily on teacher-student architectures, which may introduce additional training complexity.

While these studies have made significant progress in GAN optimization, HEVC compression, and energy efficiency, none have explicitly explored the synergy of Knowledge Distillation with GAN-based HEVC compression for energy-efficient video transmission.

Our work builds on these advancements by proposing a GAN-based HEVC framework that integrates Knowledge Distillation to achieve up to a 10% reduction in energy consumption while preserving high perceptual quality. Unlike previous works we focus on both inference efficiency and compression performance, ensuring our approach is scalable for real-world applications.

3. BACKGROUND: GENERATIVE ADVERSARIAL NETWORK AND VIDEO COMPRESSION

Generative Adversarial Networks (GANs) are machine learning models designed to generate data similar to the training datasets. GANs are composed of two neural networks: the generator and the discriminator. The generator creates fake data, while the discriminator attempts to distinguish it. The following figure illustrates the principle of GANs.

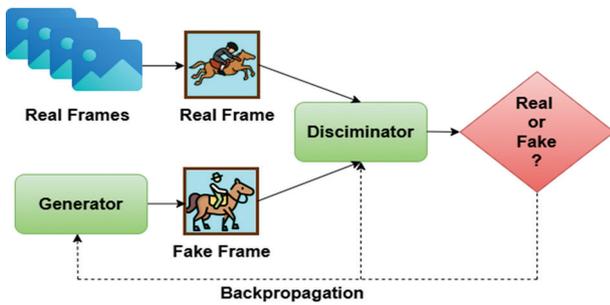


Fig. 1. GANs principle

The Generator G generates synthetic data, such as images or video frames, based on a random noise vector z . Its main goal is to produce data that resembles the training dataset. The Discriminator D evaluates the data by distinguishing between real samples and fake ones. Consequently, it provides feedback to improve the generator's performance [17]. The adversarial training process is driven by a minimax optimization problem as presented in equation (1).

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

Where $p_{data}(x)$ represents the real data distribution, and $p_z(x)$ refers to the prior distribution of the latent space. The generator aims to minimize the function, by generating output indistinguishable from the real data, whereas the discriminator seeks to maximize the function, enhancing its ability to identify real data from fake data. This adversarial training reaches equilibrium when the discriminator is no longer able to differentiate real from generated data, achieving what is known as a Nash equilibrium.

GANs have demonstrated exceptional performance in various domains such as image generation, video creation, and data enhancement. In the context of video compression, GANs have been used to enhance the perceptual quality of the reconstructed video frames, effectively minimizing artifacts while preserving details. How-

ever, the computational complexity of GANs presents a significant challenge for deployment in environments with limited resources. GANs have been applied across various domains. In our prior work [18], we integrated GANs into the High Efficiency Video Coding (HEVC) encoder to enhance video compression performance, the proposed approach leveraged the generator's ability to refine compressed frames resulting in improved visual quality, as demonstrated by increased PSNR and SSIM metrics. However, this enhanced performance highlighted another concern, which is the increased energy consumption associated with GAN-based methods.

In this paper, we extend our previous work by incorporating knowledge distillation into the GAN-based HEVC framework. Knowledge distillation is a compression technique in which a smaller and more efficient student model learns to replicate the behavior of a larger and more complex teacher model [19]. In the next section, we will detail the methodology adopted to integrate knowledge distillation with GANs to achieve a sustainable and efficient video compression solution.

4. OUR METHODOLOGY

The proposed teacher-student GAN model algorithm for HEVC compression uses a pre-trained teacher model to generate outputs, which are used to train a smaller student model. Next, we process each batch of data through forward passes for both models. Subsequently, a combined loss is computed from both distillation and ground truth losses to facilitate learning. Then comes the evaluation of the student model, achieving enhanced performance in terms of video quality and energy consumption. The following figure presents the process adopted in our research.

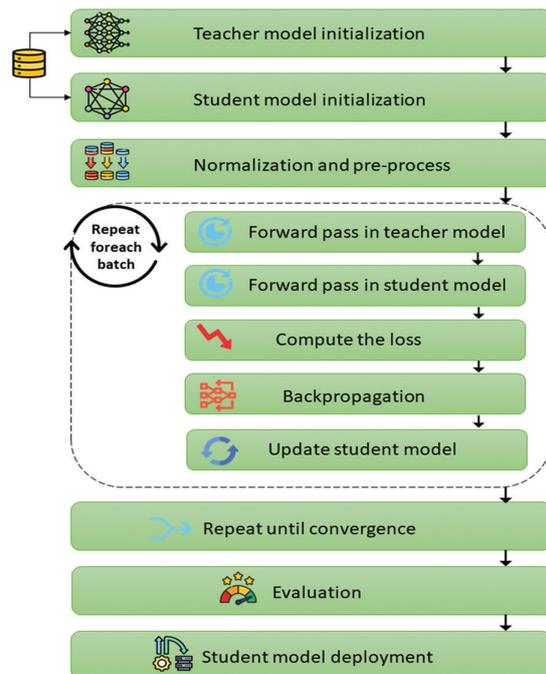


Fig. 2. GAN-based HEVC with Knowledge Distillation Methodology

In this next paragraph, we will provide a deeper understanding of our approach.

- **Step 1:** Teacher Model T initialization

During this step, we load the pre-trained teacher model T, which is a large and complex GAN trained on HEVC video frames.

- **Step 2:** Student Model S Initialization

In this step, we define the student model S, which is a smaller and more efficient GAN model.

- **Step 3:** Data preparation

The training dataset D consists of HEVC-compressed video frames, which undergo pre-processing and augmentation to improve generalization. The augmentation techniques applied include random cropping, horizontal flipping, Gaussian noise injection, and brightness adjustment. These transformations help the student model learn robust features and enhance its ability to generalize across different video sequences.

- **Step 4:** For each batch $b \in D$:

The following steps are performed for each batch of training data

- a. Forward pass in teacher model

Firstly, we pass the input $x \in b$ through the teacher model T.

Then, we compute through equation (2) the teacher's output logits z_T softened by the temperature $T_{temperature}$ [20], where $T_{temperature}$ controls the smoothness of the teacher's predictions. We empirically tuned $T_{temperature}$ in the range of 1 to 5, selecting the optimal value based on validation performance.

$$z_T = \frac{T(x)}{T_{temperature}} \quad (2)$$

- b. Forward pass in student model

In this step, we pass the same input x through the student model S.

Then, we compute the student's outputs logits presented by the equation (3):

$$z_S = S(x) \quad (3)$$

- c. Compute the loss:

In this step, we compute the distillation loss that measures the similarity between the teacher and predictions using equation (4). We use softmax with temperature for soft labels [20].

$$L_{distill} = KL(\text{softmax}(z_T) || \text{softmax}(z_S)) \quad (4)$$

After that, we compute the ground truth loss using function cross-entropy loss as shown in equation (5) [19]:

$$L = \alpha L_{true} + (1 - \alpha) L_{distill} \quad (5)$$

Finally, in equation (6) we combine both losses using weighting factor α [19]:

$$L = \alpha L_{true} + (1 - \alpha) L_{distill} \quad (6)$$

- d. Backpropagation:

Gradients are computed using the Adam optimizer, with an initial learning rate of 10^{-4} . To ensure stable training and convergence, we apply a learning rate decay of 0.1 every 20 epochs if validation loss stagnates. This prevents overfitting while maintaining optimization efficiency.

- e. Update Student Model:

Student model S is updated using the Adam optimizer with a learning rate $\eta = 10^{-4}$. This learning rate was selected based on its widespread success in stabilizing GAN training and ensuring gradual convergence while avoiding large gradient updates, as noted in [19].

- f. Convergence Criteria:

Training continues until:

- The combined loss L plateaus, with no significant improvement over 10 epochs. This threshold was chosen to ensure computational efficiency while avoiding overfitting.
- The validation metrics meet the threshold, with PSNR exceeding 40 dB and energy consumption reduced by at least 30% compared to GAN-based HEVC. These thresholds were informed by prior experiments indicating that these values balance video quality and energy savings.
- An early stopping mechanism halts training if validation loss increases for five consecutive epochs, a standard approach to mitigate overfitting.

- **Step 5:** Repeat Until Convergence:

The training process continues until one of the following conditions is met:

- The total loss L stabilizes, with no significant improvement over 10 consecutive epochs.
- PSNR exceeds 40 dB, and energy consumption is reduced by at least 30% compared to GAN-based HEVC.
- An early stopping mechanism is triggered if the validation loss increases for five consecutive epochs.

- **Step 6:** Evaluation of the student model

In this step and after the training, we evaluate the student model S on a validation set, by measuring performance in terms of:

- Video quality (PSNR).
- Energy consumption (reduced IA Energy and GT Energy).

- **Step 7:** Student Model Deployment

In this final step, we deploy the trained student model S for GAN-based HEVC, benefiting from

faster inference, fewer parameters, and reduced energy consumption.

5. PROCESS AND PERFORMANCE EVALUATION

a. Experiments:

• Implementation Framework

To implement the framework introduced in this paper, we used Python and its extensive libraries due to their robustness and flexibility. For deep learning model development and testing, we employed PyTorch, which facilitated the design of the GAN architecture and the implementation of the knowledge distillation algorithm within the teacher-student framework. NumPy and Pandas were utilized for efficient data preprocessing and handling of large datasets, ensuring seamless preparation of the HEVC video data.

• Dataset

For our experiments, we utilized the BVI-CC dataset [20], a publicly available resource designed for video compression and quality assessment research. This dataset comprises nine source video sequences, each with a native Ultra High Definition (UHD) resolution of 3840x2160 pixels, offering a diverse range of spatial-temporal characteristics. For our study, we downsampled these sequences to a full HHD resolution of 1920x1080 pixels with our experimental setup. Each sequence contains 65 frames, providing a substantial amount of data for training and evaluation purposes. The BVI-CC dataset includes videos from various genres, ensuring a comprehensive evaluation across different content types. The dataset is divided into training, validation, and testing sets, with a typical split of 70% for training, 15% for validation, and 15% for testing, facilitating effective model development and assessment. This dataset is particularly suitable for our research as it includes sequences encoded with High-Efficiency Video Coding (HEVC), aligning with our focus on HEVC video compression. The availability of both original and compressed versions of the videos in BVI-CC allows for a thorough analysis of compression performance and quality assessment, making it an ideal choice for our study.

• Knowledge distillation framework

The teacher model is a large GAN pre-trained on the HEVC dataset, while the student model is a smaller version of the same architecture. Knowledge distillation involves transferring knowledge from the teacher to the student using a temperature scaling factor of 3.0. The combined loss functions used for training the students include:

- Distillation Loss: KL divergence with softened outputs from the teacher model $\alpha=0,3$
- Ground Truth Loss: Cross-entropy loss with the original training labels ($1-\alpha=0,7$).

As for the training hyperparameters:

- Batch size: 32
- Learning rate: 10-4 (optimized using Adam Optimizer)
- Convergence criteria: Training stops when PSNR exceeds 40 dB and energy consumption reduces by at least 30%.
- Early stopping: Triggered if validation loss does not improve for 5 consecutive epochs.
- **Energy Monitoring and visualization**

For energy consumption monitoring, we used Intel Power Gadget for CPU metrics and NVIDIA System Management Interface for GPU usage and power consumption. Model performance was visualized using Matplotlib, showcasing improvements in video quality (PSNR/SSIM) and energy efficiency across all experimental setups.

This comprehensive setup enabled us to validate the efficiency of the GAN-based HEVC framework with knowledge distillation, achieving a significant reduction in energy consumption while preserving video quality.

b. Results and discussion:

Our novel approach demonstrates competitive performance in terms of reducing energy consumption while maintaining high video quality. In this work, we reduced energy consumption by up to 30% during video encoding. This optimization is most evident in the reduction of both CPU (IA Energy) and GPU (GT Energy) usage, which are critical components of the computational cost introduced by GANs. The following figure presents a comparison between, the energy consumption in Traditional HEVC, GAN-based HEVC, and Knowledge Distillation GAN-based HEVC.

In Fig. 3, firstly, the HEVC serves as a baseline and it shows a predictable behavior, and consistent increase in energy consumption as more frames are encoded. Secondly, the GAN-based HEVC demonstrates a significant rise in energy usage by 40% due to the computational costs of GANs. Finally, the Knowledge Distillation GAN-based algorithm presents a significant reduction of energy consumption of 30%, while still delivering equivalent video quality as illustrated in the next figure which demonstrates the PSNR values across all three methods.

The Graph in Fig. 4 compares the PSNR values across three methods for encoding 100 video frames.

As shown, the PSNR values for traditional HEVC fluctuate between, 41,2 dB and 43,3 dB, exhibiting the standard visual quality achievable by HEVC without any enhancement techniques. As for HEVC with GANs, it demonstrates an improvement over traditional HEVC, with PSNR values typically around 0,5 dB higher across all frames. HEVC with GANs and Knowledge Distillation

performs slightly below the GAN-based HEVC but still maintains a consistent PSNR improvement over the traditional algorithm. On average, the PSNR values are 0,3 dB higher than HEVC demonstrating that knowledge Distillation successfully retains the quality enhancement of GANs while reducing the computational complexity and the energy consumption. The minimal drop of PSNR is a minor trade-off to the gain in energy efficiency which makes GAN-bases HEVC with Knowledge Distillation an attractive solution for large-scale and energy-constrained video applications. The following figure presents a comparison of 5 frames from the original video, HEVC, HEVC with GANs, and HEVC with GANs and Knowledge Distillation for 5 random frames.

Fig. 5. demonstrates how HEVC with GANs enhances visual fidelity, while Knowledge Distillation further refines the balance between quality and computational efficiency. To provide a clearer illustration of the com-

parison, the next figure represents a detailed analysis of a single frame, showcasing the difference between the three algorithms. The below above shows the effects of three different encoding methods. The original frame presents a reference point, showing the unmodified version. The HEVC encoded frame demonstrates the results of traditional HEVC.

While it seeks to compress the frames, some visual artifacts can be observed, indicating a compromise between compression efficiency and quality. The third frame, HEVC with GANs, illustrates the application of Generative Adversarial Networks to enhance the encoding process. It is evident that GANs enhance perceptual quality as shown by finer details and reduced artifacts compared to traditional HEVC.

The last frame highlights the improvement in encoding quality emphasizing its importance in improving compression performance.

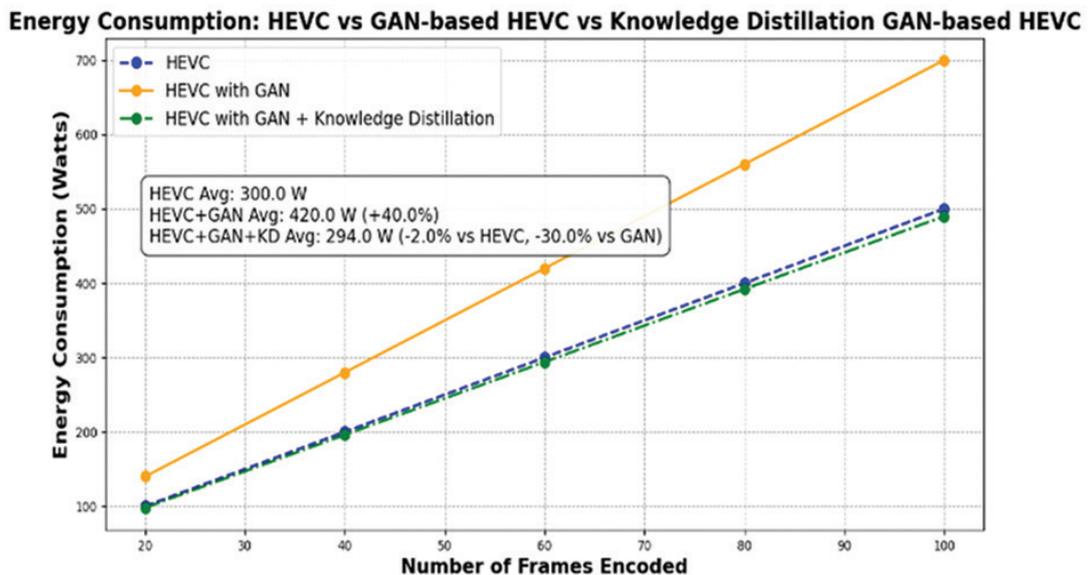


Fig. 3. Energy Consumption between HEVC, GAN-based HEVC, and Knowledge Distillation GAN-based HEVC

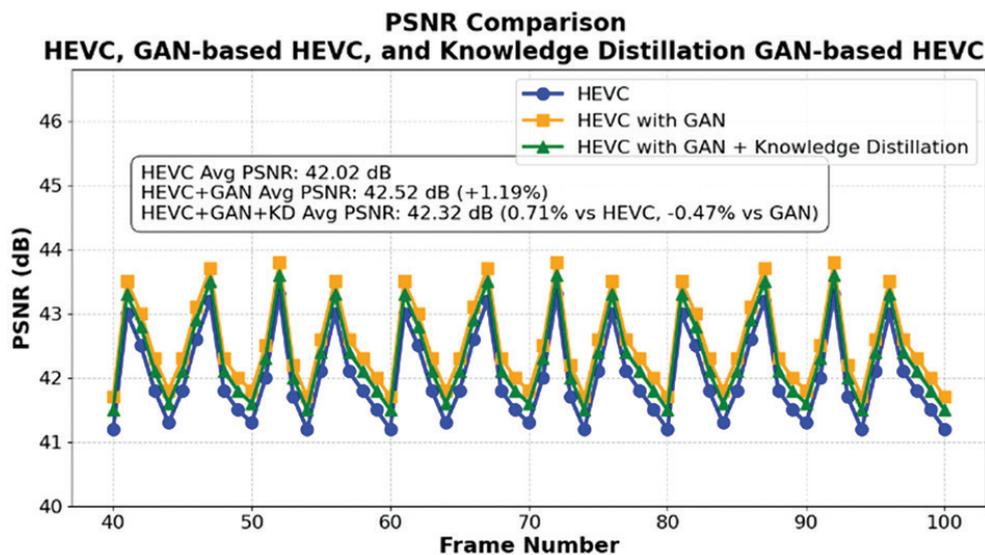


Fig. 4. PSNR values comparison between HEVC, GAN-based HEVC, and Knowledge Distillation GAN-based HEVC

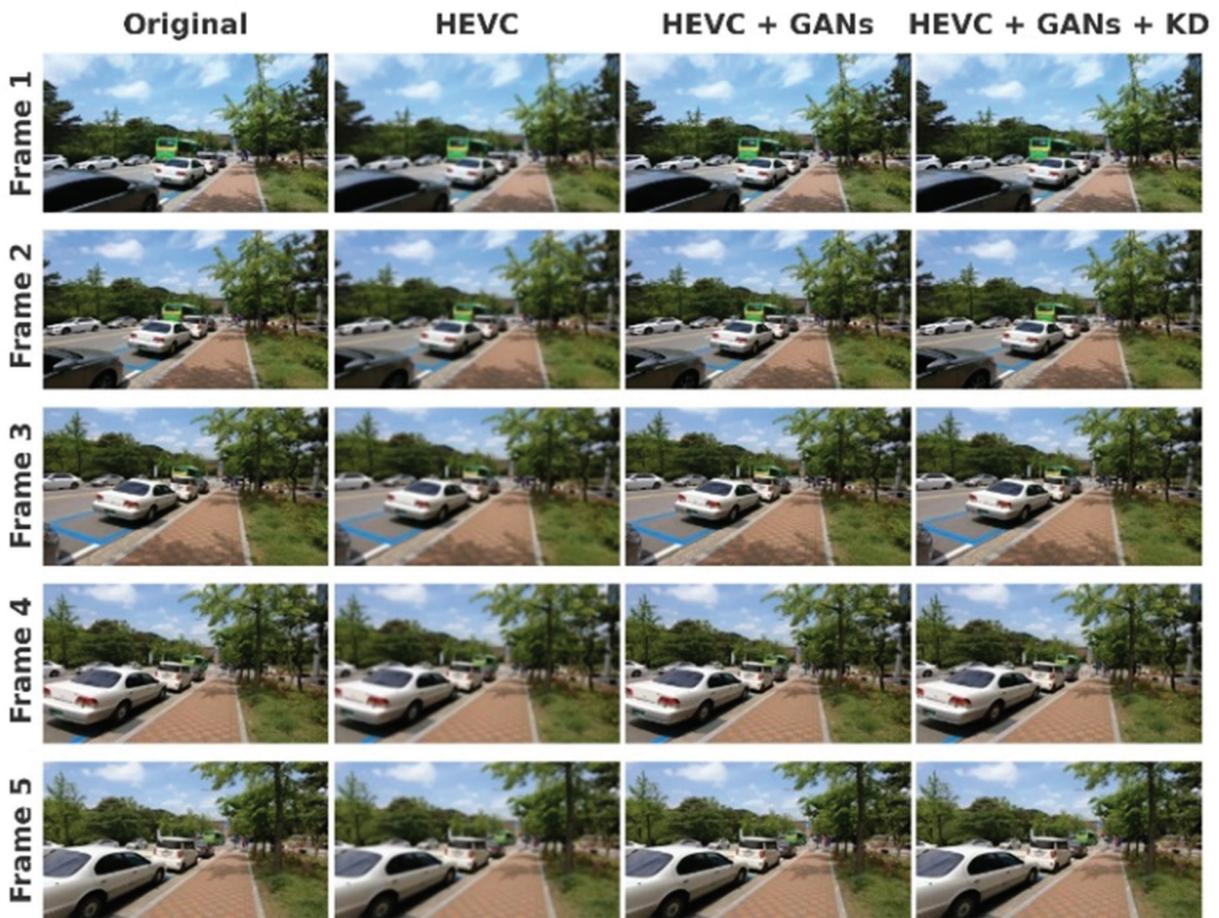


Fig. 5. Comparison between five frames of original video, HEVC, HEVC with GANs, and HEVC with GANs and Knowledge Distillation

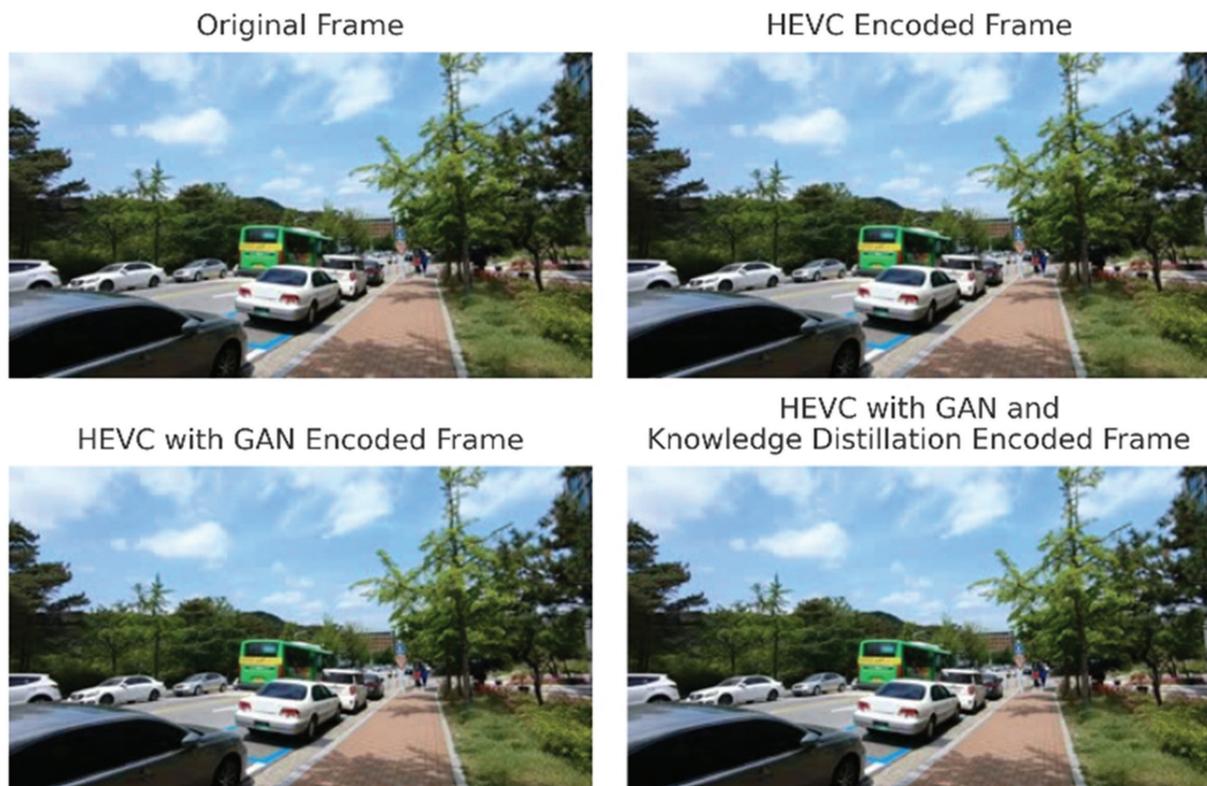


Fig. 6. Frames comparison between original video, HEVC, HEVC with GANs, and HEVC with GANs and Knowledge Distillation

Table 1. Table compares Energy consumption and quality between original video, HEVC, HEVC with GANs, and HEVC with GANs and Knowledge Distillation

Study	Energy Reduction	Video Quality
GAN-based HEVC with Knowledge Distillation (Our work)	Up to 30%	Hight PSNR improvement
SetGAN: Scale and Energy Trade-off [11]	Up to 20%	Moderate
GAN-enhanced Ensemble Model [12]	Up to 15%	Moderate (Specific cases)
Adaptive Computation in Energy-based Models [13]	Up to 18%	Varies
General Knowledge Distillation Methods [8]	Up to 25%	High

6. EFFICIENCY BENCHMARKING

Our work was benchmarked against prior research efforts focused on reducing energy consumption in video compression. The table 1. provides a comparison between various studies on energy-efficient video compression approaches. It highlights that our proposed approach achieves the highest energy reduction of up to 30% while maintaining superior video quality.

7. CONCLUSION

In this paper, we introduced a new approach that addresses the energy consumption challenge in GAN-based HEVC using knowledge distillation. Our results demonstrate that GAN-based HEVC achieves improvements in visual quality at the cost of increased energy consumption.

However, when combined with knowledge distillation, energy consumption is reduced by an impressive 30% while maintaining competitive visual quality.

Our work highlights the potential of knowledge distillation in balancing performance and energy efficiency, paving the way for more sustainable video compression techniques.

Despite these promising results, our approach has certain limitations. Future research could explore its scalability to larger datasets and diverse video resolutions, ensuring generalizability across different video formats. Additionally, further optimizations tailored to specific hardware architecture (edge devices, mobile processors, and specialized AI accelerators could enhance real-world deployment Investigating adaptive knowledge distillation techniques that dynamically adjust compression settings based on network conditions and content complexity is another promising direction.

8. REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding

(HEVC) standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 22, No. 12, 2012, pp. 1649-1668.

- [2] D. Liu, Y. Li, J. Lin, H. Li, F. Wu, "Deep learning-based video coding: A review and a case study", ACM Computing Surveys, Vol. 53, No. 1, 2020, pp. 1-35.
- [3] F. Mentzer, G. D. Toderici, M. Tschannen, E. Augustsson, "High-fidelity generative image compression", Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 11913-11924.
- [4] R. Erol, "Case Studies for Energy Efficient Machine Learning Inference Acceleration", University of Arkansas at Little Rock, AR, USA, PhD Thesis, 2024.
- [5] Y. Dong, J. Tao, Y. Zhang, W. Lin, J. Ai, "Deep learning in aircraft design, dynamics, and control: Review and prospects", IEEE Transactions on Aerospace and Electronic Systems, Vol. 57, No. 4, 2021, pp. 2346-2368.
- [6] Z. Wang, T. Luo, R. S. M. Goh, J. T. Zhou, "EDCompress: Energy-aware model compression for data-flows", IEEE Transactions on Neural Networks and Learning Systems, Vol. 35, No. 1, 2022, pp. 208-220.
- [7] G. Hinton, "Distilling the Knowledge in a Neural Network", arXiv:1503.02531, 2015.
- [8] J. Gou, B. Yu, S. J. Maybank, D. Tao, "Knowledge distillation: A survey", International Journal of Computer Vision, Vol. 129, No. 6, 2021, pp. 1789-1819.
- [9] D. Saxena, J. Cao, J. Xu, T. Kulshrestha, "RG-GAN: Dynamic Regenerative Pruning for Data-Efficient Generative Adversarial Networks", Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20-27 February 2024, pp. 4704-4712.
- [10] D. Saxena, J. Cao, J. Xu, T. Kulshrestha, "Re-GAN: Data-efficient GANs training via architectural re-configuration", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17-24 June 2023, pp. 16230-16240.
- [11] N. K. Jayakodi, J. R. Doppa, P. P. Pande, "SETGAN: Scale and Energy Trade-off GANs for Image Applications on Mobile Platforms", arXiv:2103.12896, 2021.

- [12] D. Wu, K. Hur, Z. Xiao, "A GAN-enhanced ensemble model for energy consumption forecasting in large commercial buildings", *IEEE Access*, Vol. 9, 2021, pp. 158820-158830.
- [13] Y. Du, I. Mordatch, "Implicit generation and modeling with energy-based models", *Advances in Neural Information Processing Systems*, Vol. 32, 2019.
- [14] G. Zhong, J. Wang, J. Hu, F. Liang, "A GAN-based video intra coding", *Electronics*, Vol. 10, No. 2, 2021, p. 132.
- [15] K. Wu, Z. Li, Y. Yang, Q. Liu, "Deep video compression based on long-range temporal context learning", *Computer Vision and Image Understanding*, Vol. 248, 2024, p. 104127.
- [16] L. Helminger, R. Azevedo, A. Djelouah, M. Gross, C. Schroers, "Microdosing: Knowledge distillation for GAN-based compression", *arXiv:2201.02624*, 2022.
- [17] I. Goodfellow et al. "Generative adversarial nets", *Advances in Neural Information Processing Systems*, Vol. 27, 2014.
- [18] H. Hardi, I. F. E. Fatani, "Optimizing Video Compression Quality Using AI-Boosted HEVC", *Proceedings of the 6th International Conference on Intelligent Computing in Data Sciences*, 2024, pp. 1-4.
- [19] I. Goodfellow et al. "Generative adversarial networks", *Communications of the ACM*, Vol. 63, No. 11, 2020, pp. 139-144.
- [20] A. Katsenou, F. Zhang, M. Afonso, G. Dimitrov, D. R. Bull, "BVI-CC: A dataset for research on video compression and quality assessment", *Frontiers in Signal Processing*, Vol. 2, 2022, p. 874200.

MLbFA: A Machine Learning-Based Face Anti-Spoofing Detection Framework under Replay Attack

Original Scientific Paper

Vijay V. Chakole*

G H Rasoni University,
Amravati, Maharashtra, India
chakole.vijay@gmail.com

Dr. Swati R. Dixit

G H Rasoni College of Engineering and Management,
Nagpur, Maharashtra, India
swati.dixit@raisoni.net

*Corresponding author

Abstract – The primary aim of the research paper is to deploy an efficient automated face anti-spoofing system that could consider replay attacks in the presence of partial occlusions. For this purpose, the article introduces a novel machine learning-based face-anti-spoofing (MLbFA) framework. The system incorporates a modified version of the difference of the Gaussian technique to compute the overall contrast of the input images which is later used to enhance the contrast of the image using contrast correction. On the other hand, the image details, especially the edges are enhanced for significant feature contribution using a Beltrami filter. The contrast-cured and extremity-enhanced images are averaged to obtain a finer image. Face cropping is achieved using the Bounding-Box algorithm to reduce computational complexity and improve classification accuracy for region-bounded feature extraction. Quality conventional or handcrafted features (CF/HF) are extracted through various descriptors from the region of interest (ROI). The features are reduced in dimension using principal component analysis (PCA) and portioned in training and testing sets with a 75%:25% ratio respectively. An experimental study showed that the proposed MLbFA model using a Support Vector Machine (SVM) outperforms other recent existing face anti-spoofing competing techniques with an improvement of 0.11% compared to the best-performing Edge-Net Autoencoder model concerning the classification accuracy.

Keywords: Face anti-spoofing, machine learning, Difference of Gaussian, Beltrami filter, Bounding-box algorithm, conventional features, PCA, and SVM

Received: September 30, 2024; Received in revised form: December 23, 2024; Accepted: December 23, 2024

1. INTRODUCTION

Biometric authentication can be primarily classified as possession-based and knowledge-based authentication. Knowledge-based is secretly available with the subject that is only known to the subject himself. However, biometric authentication models are vulnerable to spoof threats where an intruder or fraudster endeavors to steal or compromise it. The type of attack can be differentiated depending on the biometric modality whether the system uses the subject iris pattern, his fingerprints, his facial features, audio, or a keystroke. Despite this, few traits are non-comparable. Therefore, a specifically designed algorithm is required to identify the spoof attacks since each biometric trait

has its merits and limitations. In the case of spoof attacks, the fraudster peeps through the vulnerabilities in the biometric system. The fraudster enters the secured biometric-based system attempting to deceive another person. This leads to the significant requirement of an anti-spoofing technique for securing biometric systems and preventing unauthorized trait replicas.

Despite the advancements being made in the field of biometrics, various authentication systems still fail miserably while detecting a spoofing attack. This leads to a positive response being made to an unauthorized individual thereby giving him the overall control of the system unknowingly. Many well-known organizations and industries have already fallen into this trap of breach and

spoofing attacks. Hence, a well-developed automated system that can recognize the same in real time has become mandatory. For this purpose, various researchers have used many machine and deep learning classifiers, who tend to extract relevant features from the dataset and later pre-process them for better accuracy. However, due to data imbalance, certain interference peeping in CCD and CMOS sensors, uneven illuminations, distortions, and partial occlusions due to hair, specs, scarfs, organs, etc. the anti-spoofing systems have suffered setbacks in classifying the spoofed and real images.

The work proposed by authors [1] tends to automate a robust system by making use of a Laplacian filter that not only filters the relevant images but also enhances the overall output thus generated. On the other hand authors [2] enlightened the implementation of Schmid as the filter which could uplift the disadvantages of the image and convert them into a readable format. Meanwhile, authors [3] proposed to modify the existing working of the DOG filter by adding an extra layer that could destroy the noise and generate high-frequency edges. A correction algorithm was thus used and further pre-processing of the images was done by eliminating the interference of light. Even though various pre-processing techniques were primarily used, authors [4] executed certain anti-spoofing techniques that could easily differentiate between original and fake images.

In addition to this, the generated 2D-based spoofed images were exposed to various forms of noise distortions that included glossy photo papers and digital screens. Color distortions were also faced in the same process which occurred as a result of resolution from the screen. Apart from noise and color distortions, distortions in the form of facial deformations also occurred which were proposed by authors [5]. Hence, the concept of landmark components was brought into the picture by the author [6-7] who finally segmented the image into equal halves so that optimized accuracy could be obtained.

The techniques [8, 9] incorporating textural properties of the spoofed and authentic images lacked generalization ability even though they showed rapidness in response. Models based on extracting color depth information from a 3-dimensional face using multiple image frames [10, 11] required higher reliability to classify the 2D spoof faces. On the other hand, losing low-frequency details while highlighting the higher-frequency components was associated with the process of recapturing video replays or printed photos. These models were successful and showed better sensitivity when the intra-dataset test was concerned, however, for improved performance and inter-dataset tests, diverse multiscale features were required.

Deciding on a perfect spoof mechanism out of the existing approaches would need studying the features, advantages, and negative remarks of each approach. With this in mind, facial biometrics can be seen as a special case, since multimodality can take advantage of multiple facial properties (e.g., texture, shape, and tempera-

ture) to avoid spoof attacks. Spoofing attacks persist be a security challenge for face biometric systems, and there was much effort in the field to find robust methods. However, all these efforts have been following the same recipe, not favouring breakthroughs in the field. Many works of face spoofing detection emphasize 2D attacks by presenting printed photos or replaying recorded videos, and 3D attacks have been recently studied due to the technological advancements in 3D printers and reconstruction. Although perfect results on public data sets have been achieved by many works, there is a considerable gap in moving from academic research to real-world applications in an effective way.

The paper's contributions can be summarized as:

- The ROI is extracted so that the system can eliminate distortions and focus on the facial image by creating a boundary box along the edges of the image. This reduces the overall run-time complexity of the classifier and thereby increases the probability of classification accuracy of the model.
- Sufficient informative handcrafted features are extracted from the ROI that represented the faces more accurately and significantly.
- The feature dimensionality was reduced so that the obtained dataset features are minimized and only the relevant features are executed in the training and testing phase.
- The proposed MLbFA model is simpler and more efficient to discriminate real and unauthentic faces from dataset test samples and real-time test samples.

2. RELATED WORK

A generative method of probabilistic voting was introduced by authors [12] wherein they made use of an ensemble classifier along with a discrete wavelength filter. Their study aimed to perform segmentation of the face image and further calibrate it using face alignment so that the overall frequency of the residual image could be reduced. The residue from this image was later converted into a YCbCr model and texture features were extracted from the same by making use of a texture descriptor. The repository used for the implementation of the same comprised four datasets. In a similar work performed by authors [13] they extended their research work by making use of SURF features which led to the conversion of grayscale images to color images. For this purpose, the author used specific color bands for each image and further concatenated them with the SURF features. In the next stage, the overall complexity of the system was reduced by using the PCA algorithm and Fisher as the Vector. All the descriptors used for implementation were obtained using HAAR as the wavelet function comprised of 4x4 blocks. In the final stage, the sub-regions around the image were concatenated by using HSV as the implementation algorithm which eventually resulted in the overall feature vector to reduce the dimension to 64.

In another work authors [14] implemented the techniques of chromatic illumination so that the objective of differentiating between a real and forged image could be made. They proposed to extract inter-channel chromatic occurrence to obtain chromatic texture features. Softmax was used as the classifier along with LBP as the feature extractor. Combinations of 4 datasets were used (MSU, MFSD, CASIA, and FASD) and the overall discriminating factor of the system model was improvised. Similar work was offered by authors [15] wherein they made use of an ensemble learning methodology which was eventually used to analyze the existing chromatic discrepancies. The dataset however consisted of imbalanced images. In another work suggested by authors [16] they made use of SVM as the fuzzy classifier which eventually analyzed the acquired dataset from different angles and perspectives. All the images from the dataset underwent the process of feature extraction by making use of the HOG feature extractor. In addition to this, LPQ features were also simultaneously used which minimized the invariance of images that occurred.

Abhishek Mittal *et al.* [17] obtained a 10% improvement for accurately distinguishing the natural and unnatural faces using features based on the Gray level co-occurrence matrix (GLCM). The texture features-based hybrid approach was evaluated using an integration of 3 different ML classifiers.

Mays Alshaihl *et al.* [18] suggested an attention module to consider the relevant features only using the spatial features and the color depth features. Their dual module approach was efficient in seeking context-based significant features that helped their framework in improving the classification accuracy. On the other hand, their objective to improve the generalization ability of the deep learned (DL) classifier was also enhanced. The integrated model incorporating the attention module and DL network succeeded in aiding the overall performance. The authors suggested that their framework can be implemented for pixel-based attention mechanisms including quality assessments, segmentation, and face detection jobs.

Spoof strikes were detected by Junwei Zhou *et al.* [19] utilizing an inventive approach that combined LDN-TOP representation and Pro CRC (probabilistic collaborative representation-based classifier) classification pipeline. The LDN and a derivative Gaussian mask were used to learn the texture patterns of the concerned region under disturbances caused by illuminations. While LDN was broadened to spatial-temporal variant to occupy the motion features. The Pro CRC was made to learn from the LDN-TOP represented features extracted from depth images. The experimental evaluation was carried out on 3-different sets of dataset images concerning the EER and HTER (half total error rate). The authors obtained 0.37% EER on the CASIA dataset and 5.7% HTER on the UVAD dataset using a sequence-based protocol. They also carried out a time-window length analysis and demonstrated that improved out-

comes can be obtained using larger lengths at the cost of sufficient video frames. Competitive results were also shown for replay attacks in their article.

The authors [20] focused on pre-processing the images before they were submitted to a convolutional neural network constructed with 12 layers. The pre-processing concentrated on proper and precise face alignment, extracting the concerned region, and controlling the uneven illumination from face regions. Cropping was accomplished using a bounding box algorithm while the latter part dealt with three cascaded networks. The P, R, and O-net cascaded network structure controlled the illumination of the bright regions of the face images obtained from the CASIA-FASD dataset. Through experimentation, the author showed that their model attained an HTER of 1.02% and improved the classification accuracy.

The immense success of pre-trained networks in several applications was noticed by authors in [21]. They used 2-different color spaces of the input images and extracted face embeddings which were jointly used for classification using the VGG network. For better face representation, the images were denoised priority and then converted to other color spaces. The joint embeddings from CIE LUV and YCbCr color spaces obtained a false rejection ratio of 0.3% and an acceptance ratio of 0.4% approximately with an accuracy and specificity above 99%.

The LBP and SVM integrated classification approach used by authors in [22] considered a rotation invariant scheme over photo imposter dataset images. The illumination variations introduced in the dataset images were distilled using a modified Difference of Gaussian filter and filtered images were represented employing the LBP operator. Comparison with the normal combination of LBP+SVM and LBPV+SVM showed that the latter outperformed the former, however, the approach failed to perform in the case of cross-dataset samples.

Multi-channel images from the PAD dataset were part of the work introduced in [23] that are prone to various Presentation Attacks. The author utilized all the aspects of multichannel images including the depth, NIR, and color channels, and presented a PAD protocol using a combination of autoencoder and Multi-Layer Perceptron. The challenges due to Presentation Attacks were handled using a dual path framework by the authors who claimed that the individual faces represented higher disparities compared to that of complete faces.

Several researchers over the globe have contributed to developing antispoofing schemes using conventional features by different classifiers. They have used different dataset images with different levels of complexity and analyzed the performance of their scheme on single and cross-domain datasets. However, few of them lacked the generalization ability when cross-dataset or real samples were tested while others failed to classify the samples accurately. Also, in most cases, the classifier models were complex and required a longer time for training.

3. MATERIALS AND METHOD

The system thus proposed in the research paper is divided into five main stages. The implementation begins at the pre-processing stage followed by extraction of relevant features. A process of dimensionality reduction also takes place before the classification. The research finally comes to an end by evaluating and analyzing the results.

Initially, a dataset is obtained which comprises real and fake images. The first stage of pre-processing is done on these images which eventually comprises two primary techniques in parallel used to correct the image thus obtained. The technique involves contrast correction of the image and then using a filter to preserve the edges of the image. The resulting output from these techniques is further given as an input to the next stage which appears to be the feature extraction process. This happens to be the second stage of implementation and is mandatory to enhance the edges of the image and further extract only the relevant features so that the final distinguishing between the real and the fake ones can be made. A total of 2056 image features are extracted and further reduced

in terms of dimensions using the PCA in the third stage. The fourth stage of system implementation is characterized using the classification technique. For this purpose, we have used SVM as the classifier, and the dataset is fed to the training and testing phase so that evaluation of the model can be made. A percentage ratio of 75 and 25 is made respectively for training and testing purposes. Finally, the system is evaluated using performance metrics.

3.1. THE DATASET

The IDIAP Replay attack dataset is used for the implementation of the proposed research. It comprises 1300 videos including the video and photo attack. The videos lasting for 9 seconds were converted to image frames for all 50 subjects. The color images obtained bear a resolution of 320x240 dimension where the first dimension corresponds to width and the latter the height. The original sampling rate was 25 Hz and the videos were captured in two environments. The controlled environment uses homogeneous backgrounds, and curtain-covered windows in the background with good surrounding illumination.

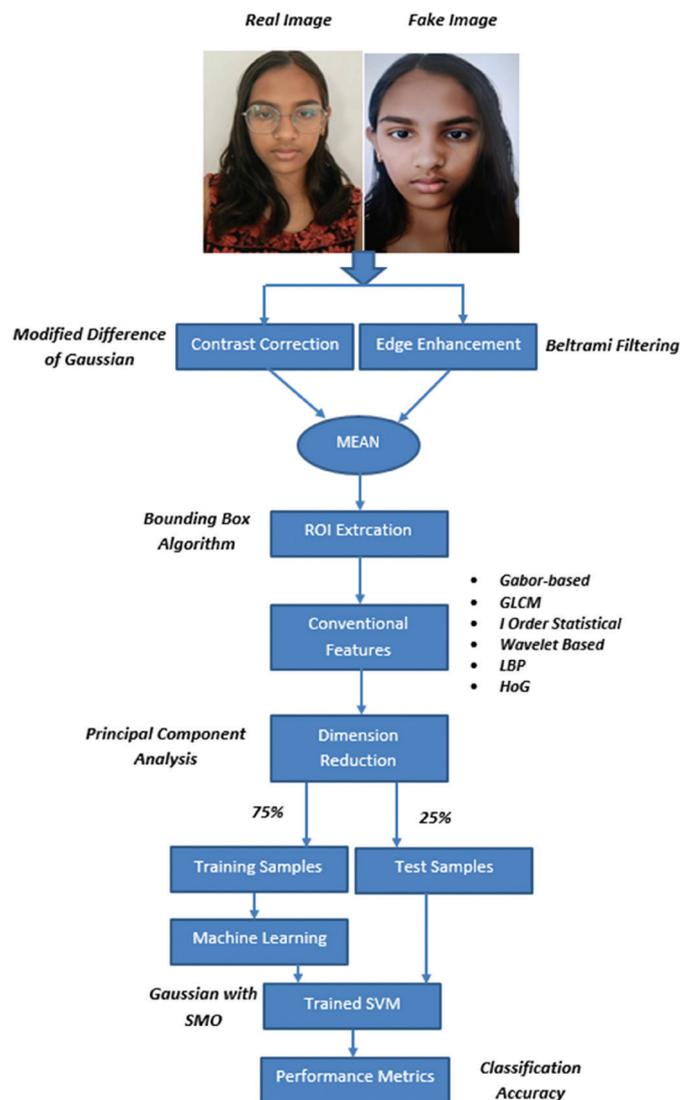


Fig. 1. The proposed MLbFA Model

The adverse environment on the other hand used a complex background with ill-illuminated surroundings and raised curtains over the windows. Ten attack video attacks for each were performed with the subject's fixed biometry and biometry from a device [8, 24].

We constructed our dataset from the above available images. We chose real images belonging to 80 subjects and fake images that included 199 subjects. The number of images for each subject from real and fake categories that were included for experimentation was 50. This was intentionally done to construct an imbalanced dataset. Therefore, images belonging to the real images class were 4000, and the fake images were 9950. Further real and fake images from real-time videos captured using Samsung Galaxy F34 with a 50 MP Camera were used to test the robustness of the proposed system and evaluate cross-test performance. The input images shown in Fig. 1 are real-time images where the left image belongs to the real category while the right belongs to the attack class.

Column 1 and column 3 of Fig. 2 show the adverse condition for real images while column 2 images are acquired in a controlled environment. Similarly, the upper row of Fig. 3 shows images from adverse environments while the bottom row depicts images obtained from the controlled environment under attack conditions.



Fig. 2. Sample of Real images from the IDIAP dataset. Column 1 and column 3 show the adverse condition (uneven illumination) for real images while column 2 images are acquired in a controlled environment

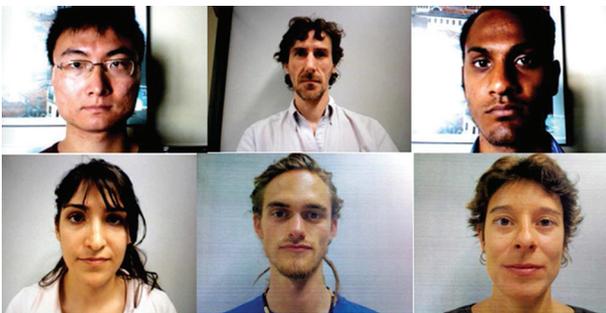


Fig. 3. Sample of spoofed images from the IDIAP dataset. Upper row of Fig. 3 shows images from adverse environments (uneven illumination) while the bottom row depicts images obtained from the controlled environment under attack conditions

3.2. CONTRAST MEASUREMENT AND CORRECTION

Thus the dataset has dual complexity issues: real and fake images are not balanced and they have discerning illuminations and backgrounds. In addition to this, the edges over the image also appear to be blurred which can affect the quality features of the image and makes it less significant during the feature extraction process. Due to this reason, the conventional features will probably fail to absorb latent details from the ROI and thus affect adversely the classification.

The image thus obtained from the dataset is heavily dependent on the spatial arrangements, edge measurements, and light illuminations in which the image is captured. Apart from this, the color of the image, resolution, and pixel size of the image also have an overall impact. To overcome this issue, we used contrast correction and edge preservation both independently on the input image. The perceived contrast is measured by computing the difference between the extremely low and extremely high-intensity pixels in the image [1]. Initially, the overall image contrast is measured using the Modified-DoG method suggested by Tadmor and Tolhurst [2], and based on the measured contrast value (equation 1) the image contrast is corrected (equation 1-8) to improve the overall quality of the image. In parallel, we also carried out distilling the image using the Beltrami Filter for edge enhancement. This filter is responsible for preserving the image edges which are distorted due to illuminations and/or background noise. Both techniques are however carried out to enhance the quality of the overall image and obtain good classification results.

Fig. 4 shows the output of modified DoG filtering. The contrast 'Cm' of an image is measured using equation (1).

$$C_m(xx, yy) = \frac{R_c(xx, yy) - R_s(xx, yy)}{R_c(xx, yy) + R_s(xx, yy)} \quad (1)$$

Where the output of the central component is,

$$R_c(xx, yy) = \sum_{i=xx-3r_c}^{i=xx+3r_c} \sum_{j=yy-3r_c}^{j=yy+3r_c} C(i - xx, j - yy) I(i, j) \quad (2)$$

While the output of the surround component is,

$$R_s(xx, yy) = \sum_{i=xx-3r_c}^{i=xx+3r_c} \sum_{j=yy-3r_c}^{j=yy+3r_c} S(i - xx, j - yy) I(i, j) \quad (3)$$

The center and surround components of the receptive field are given by, *Center component*,

$$C(xx, yy) = \exp \left[- \left(\frac{xx}{r_c} \right) \left(\frac{xx}{r_c} \right) - \left(\frac{yy}{r_c} \right) \left(\frac{yy}{r_c} \right) \right] \quad (4)$$

(xx, yy) is the spatial coordinates of the receptive field, and rc is the radius at which the sensitivity decreases to 1/e w. r. t. the peak level.

$$\text{Surround component, } S(xx, yy) = 0.85 \left(\frac{r_c}{r_s} \right) \exp \left[- \left(\frac{xx}{r_s} \right)^2 - \left(\frac{yy}{r_s} \right)^2 \right] \quad (5)$$

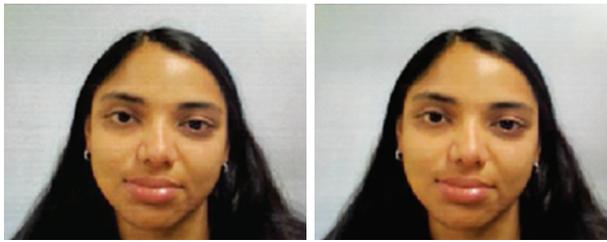


Fig. 4. Contrast measurement. (a) The original image and (b) Image obtained from the patch level contrast values

3.3. CONTRAST CORRECTION

Once the contrast measure is estimated over the image by averaging the patch level contrast values, the overall contrast value is used to enhance the image quality using expressions 6 to 8. This henceforth diminishes the processing time for feature extraction as well. 'g' represents the contrast-corrected image.

$$mm = 255 * C_m \quad (6)$$

$$f = 259 * \frac{(mm+255)}{(255*(259-mm))} \quad (7)$$

$$g = (f * (\text{Image} - 128)) + 128 \quad (8)$$

Where 'mm' and 'f' are intermediate results obtained using the contrast C_m computed using equation (1). 'Image' here represents the original image which is considered for contrast correction.

3.4. EDGE PRESERVING AND ENHANCING FILTER - BELTRAMI FILTER

The conceptual working theory of the Beltrami filter was introduced by [25]. This was majorly done to preserve the image edges from losing data due to the presence of noise. The technique makes use of a denoising filter and applies the same to the 2D images. The Beltrami filter is however capable of discarding any aliases present in the process and enhancing all the weak textures of the image while preserving the edges. The filter also makes use of various color channels which tends to separate the input image. In the proposed research we have used 20 iterations of the filter with a time step of 0.5.

The original input image and the peak signal-to-noise ratio between the filtered image, contrast-corrected image, and the mean image are shown in Figs. 5 and 6.

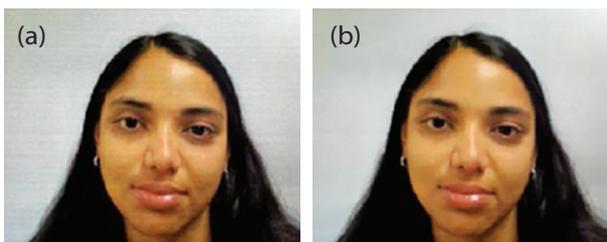


Fig. 5. (a) Original color image and (b) output of Beltrami filter with PSNR between them

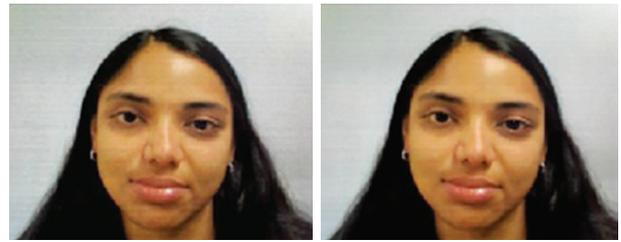
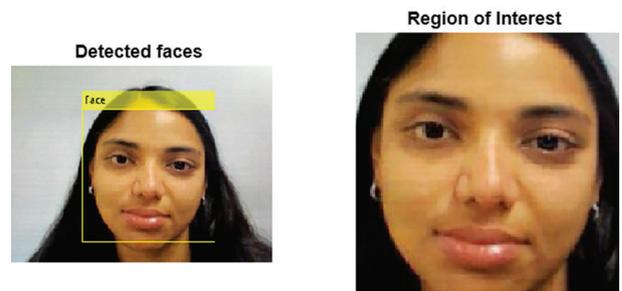


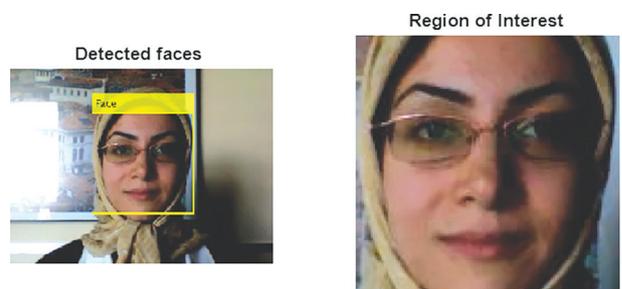
Fig. 6. (a) Output of Contrast Correction and (b) Final mean image with PSNR related to the original image

The region of the face was further extracted using the bounding box algorithm available with MATLAB 2019b. In the next stage, the extracted region of the face (ROI) is further resized to a fixed dimension [120 120 3] such that there is a minimum loss of features. This is done to ensure that different faces when localized using the bounding box algorithm carry different widths and heights. Extensive experiments were carried out on images from the dataset to set suitable dimensions for all cropped faces. The dimension chosen was relative to the smallest and the largest face detected using the bounding box algorithm. Setting smaller dimensions would cause a significant loss of features from larger cropped faces. While setting larger dimensions would probably distort the features of smaller faces. Fig. 7 below depicts the face regions automatically detected and then resized to a common dimension [120 120 3].



(a)

(a) Result of bounding box algorithm on fake image: The bounding box and cropped region



(b)

(b) Result of bounding box algorithm on real image: The bounding box and cropped region

Fig. 7. Result of Bounding Box Algorithm on samples from both classes. (a) Fake and (b) Real

3.5. FACE REPRESENTATION USING THE CONVENTIONAL FEATURES

- Gabor Filter-Based Features

The concept of Gabor features was developed by [26] and eventually used to set the obtained image representation with default filter parameters (scales = 5, orientations = 9). The number of rows and columns to be set to 39. The rows and columns were reduced by a factor of 39 through down-sampling. The final resized image was converted to a grayscale image wherein a total of 640 features were functionally extracted for representation.

- Gray Level Co-occurrence Matrix (GLCM) Based Features

The GLCM-based feature extraction technique is primarily used to extract features from the grayscale images by taking into consideration the associated properties such as contrast, homogeneity, energy, and correlation. The representation due to each of the GLCM factors was obtained in all 8 directions. Though features are obtained considering 3600 orientation (8 directions), they were averaged and the GLCM properties were represented by a single value. Thus, overall 4 values corresponding to four factors were considered for evaluation.

- Statistical Features of First-Order

A total of five features were extracted from the cropped and resized grayscale image (I_{face}). The converted images were labeled as I_{face} . The five features however included the calculation of mean, variance, standard deviation, skewness, and kurtosis which were based on probabilities obtained over I_{face} pixel intensities. The pixel intensities ' L ' are assumed to be in the range of [0 255]. The probability ' P ' is initially computed using expression (11) and eventually, all the five parameters are computed using expressions from 12 to 16.

$$L = 0:255 \quad (10)$$

Probability $P_b(x)$ of each pixel ' x ' in the image:

$$P_b(x=1 \text{ to } 255) = \frac{1}{M*N} \sum_{i=1, j=1}^{i=M, j=N} (I_{face}(i, j) = x) \quad (11)$$

$$\text{Mean, } M = \sum L .* P_b \quad (12)$$

$$\text{Variance, } V = \sum [(L - M)^2] .* P_b \quad (13)$$

$$\text{Standard deviation, } Std = \sqrt{V} \quad (14)$$

$$\text{Skewness, } S = \frac{\sum [(L - M)^3] .* P_b}{Std^3} \quad (15)$$

$$\text{Kurtosis, } K = \frac{\sum [(L - M)^4] .* P_b}{Std^4} \quad (16)$$

- Features through Wavelet transform

Wavelet-based features are constructed using 1-level decomposition over six different mother wavelets. Experiments showed that vertical and diagonal component at level 1 shows discriminative features as compared to other components. We computed the energy

and magnitude of the components for all six mother wavelets and enhanced the feature set. The six mother wavelets used for decomposition for the I_{face} image include haar, bior, debauchees and symlet (bior 3.1, bior 3.5, and bior 3.7), debauchees 3 (db3), symlet 3 (sym3), and haar).

The magnitude M_w and energy value E_w over both components is computed using the following expressions (11) and (12).

$$M_w = \frac{1}{m*n} (\sum_{r=1}^m \sum_{c=1}^n W_{xx}) \quad (17)$$

$$E_w = \frac{1}{m*n} (\sum_{r=1}^m \sum_{c=1}^n abs(W_{xx}^2)) \quad (18)$$

Where m and n are the wavelet components' row size and column size. W_{xx} is either W_{vert} or W_{diag} and represents the vertical and diagonal components.

- Histogram-Based Color Depth Features

The I_{face} color image is converted to Lab color space for color-based depth features. Histograms using 16 intensity levels are obtained over each channel from both color spaces. All six histograms are normalized using the size of the I_{face} image. The histogram values of independent channels are averaged to obtain a single histogram for both color spaces. Finally, the 16-level histograms are concatenated to obtain a 32-level feature vector to contribute to the color description. The following expressions (19) and (20) are used to compute the histograms:

In the case of RGB color space,

$$h(x \in R, G, B) = \frac{1}{M*N} h(I_{face}(x)) \text{ (16 bins)} \quad (19)$$

In the case of Lab color space,

$$h(y \in L, a, b) = \frac{1}{M*N} h(I_{face}(y)) \text{ (16 bins)} \quad (20)$$

Here ' h ' denotes the histogram.

- Haar-Wavelet-Based LBP Features

The usage of LBP as the feature extraction method enables the extractor to capture all the details of the image such as the edges, the illumination on the image, and the textural patterns. LBP is initially applied on all color channels the original image I_{face} and LBP feature are extracted. The LBP-featured channels are then used to compute the overall mean. Further, using the haar mother wavelet, the I_{face} image is wavelet decomposed to 4 levels. Similar LBP features are obtained on all four wavelet tributaries and averaged at the end.

The final haar-wavelet-based LBP features are obtained by averaging the mean LBP of the color image and mean LBP features obtained at different levels as shown in Fig. 8. Implementing this step ensures that image details are not lost and the system can eventually differentiate between a real image and a fake one. The figure below illustrates the architectural mechanism for the process of feature extraction using LBP.

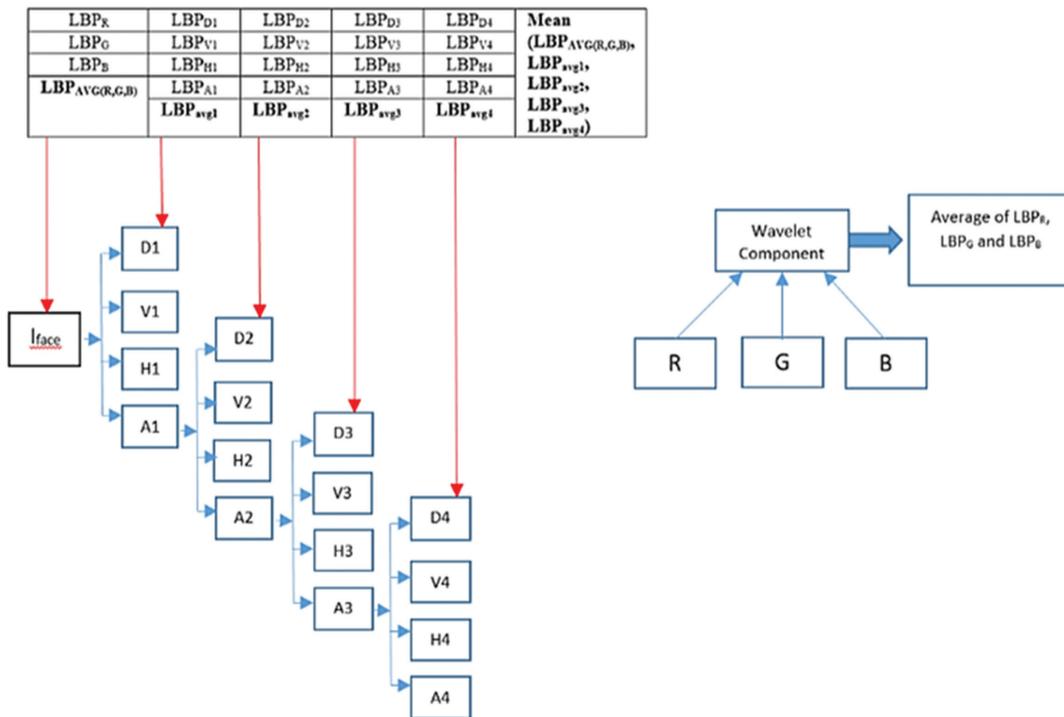


Fig. 8. Haar-wavelet-based LBP feature extraction

• Histogram of Gradient (HoG)-Based Features

HoG-based features are extracted from the Iface image and all its channels (R, G, and B) using a cell size of [16 16]. Later the HoG features were averaged to obtain the final set of features with dimension 1296. The description of conventional features with their respective dimensions is depicted in Table 1 below.

Table 1. Conventional features and respective dimension

Descriptors	Type of <i>I_{face}</i> image	Dimension
HoG	color	1296
Gabor	grayscale	640
Haar-Wavelet-Based LBP	color	59
RGB & Lab color space-based Histograms	color	32
Wavelet-based	grayscale	24
1 order Statistical	grayscale	5
GLCM	grayscale	4

4. RESULTS AND DISCUSSIONS

The performance metric that is used to evaluate the performance of the proposed Spoof detection framework is classification accuracy. It is simply computed by calculating the ratio of samples correctly classified to total samples. It is expressed by the following expression (21):

$$\text{Accuracy (\%)} = \frac{\text{Number of samples correctly classified}}{\text{Total test samples}} \times 100 \quad (21)$$

Experiments conducted on 4000 (80x50) authentic set images and 9950 (199x50) spoofed images showed that the machine learning-based SVM classifier using the conventional features achieved 99.89% training accuracy. We selected 75% of samples from each sub-

ject from both categories for training the SVM with a Gaussian kernel and 'SMO' solver. The remaining 25% of samples from each category were used for assessment which were selected randomly each time. The SVM was trained and tested 50 times and the mean accuracy was considered. The targets assigned to both classes were 0 and 1 respectively for fake and authentic samples. As seen from Table 1, more than 2000 features with different descriptors were used to represent a single image from the dataset.

The efficacy of the conventional features can be seen from the performance of the proposed MLbFA model which is shown in Table 2. We compared the proposed MLbFA model performance with other state-of-the-art competing models utilizing a similar type of dataset involving a Replay and presentation attack. As seen from Table 2, several feature-based, machine learning-based approaches and deep learning-based techniques are used by researchers to differentiate the real and spoofed classes. Though they have used different datasets and volumes of samples, the objective is to devise a solution that can classify real images from unauthentic faces.

Agarwal A. et al. [27] used a simple approach to verify the two-class samples based on the weighted sum over SVM fusion obtained from the Haralick features. The weighted sum rule fusion model was used over the conventional features obtained on the color channels of the original and the face-detected image. The authors subjected the color channels to redundant discrete wavelet transform and extracted Haralick features. The features were combined using SVM fusion and based on the weighted sum the decision was considered.

The author obtained a remarkable accuracy of 99.08% (error=0.92%). The authors in [36] utilized an RGCS-ConvNeXt using a convolutional neural network and obtained a significant accuracy of 99.25% whereas the hybrid network CNN-VGG16 introduced in [38] outperformed all previous techniques improving the classification to 99.50%. The author used three different approaches and obtained the best results using the CNN-VGG16-based features and classifying them with machine learning. The features were obtained using the HSV and YUV color space formats. The best performance for this model is limited to the scarcity of datasets required for deep-learned models.

The authors claimed that handcrafted features are unable to obtain sufficient representation for the images owing to satisfactory performance. The superior result using our proposed MLbFA model thus verifies that the features representing the dataset images for spoof detection are efficient and informative. We outperform all other anti-spoofing techniques with a simpler but efficient spoof detection mechanism. The simplicity is regarding the use of conventional features and SVM instead of deep-learned models with complex architectures.

The confusion matrix corresponding to one of the iterations is shown in Fig. 9. The number of test samples (feature vector) for real and fake images correspondingly were 1000 and 2488 (25%). The SVM achieved 100% accuracy while training and classified the test samples with an accuracy of 99.82%. The individual accuracies are 99.7% and 99.95% respectively.

	Real	Fake	
Real	997	3	99.7
Fake	1	2487	99.95

Fig. 9. Confusion Matrix for an iteration

The reason why more real samples are affected is associated with the class imbalance. The number of samples in the fake category are higher than the real samples. Therefore, even though the SVM trained accurately using the training set, few samples which are ambiguous due to illuminations, backgrounds etc. are mostly aligned towards the higher class. In a similar way, accuracies are obtained using random test samples and the mean accuracy is computed.

The significant part of the proposed work is the pre-processing stage which involves contrast measurement and correction along with edge-preserving filter operation using the Beltrami filter and the conventional coarse (GLCM, Color, First order Statistical, and wavelet features) and fine (Gabor, HoG and LBP features) qual-

ity features. The parallel process of contrast correction and filtering and then averaging resultant images mitigated the effect of uneven illumination and uplifted the edges in the original image without much loss. The feature extraction operators were able to perform their duties independently and each of them contributed positively to obtain a robust feature set. Reducing the dimension of the features and the samples improved the performance in terms of time complexity and computational complexity of the classifier.

Table 2. Comparative analysis of the proposed MLbFA model with other competing anti-spoofing techniques

Method	Year	Dataset	Accuracy
Score fusion of Partition images [27]	2017	TIFADB	99.08%
GFA-CNN [28]	2020	Siw	95.02%
NAS-FAS [29]	2021	MSU-MFSD	95.85%
Morphological SVM [30]	2021	FPAD	97.21%
Edge-Net Autoencoder [31]	2021	IDIAP	99.87%
Deep CNN [32]	2022	IDIAP	98.21
EBDG [33]	2022	MSU-MFSD	97.17
CNN [34]	2023	IDIAP	98.36
IADG [35]	2023	MSU-MFSD	98.19
RGCS ConvNeXt [36]	2024	Siw	99.25
UCDCN [37]	2024	Replay Attack	99.18
CNN-VGG16 HSV LUV [38]	2024	NUAA Imposter	99.5
Proposed MLbFA	2024	IDIAP	99.98

In the second part, we acquired real-time videos from two real subjects and their photos. The videos were converted to frames and further partitioned into two sets. A set containing 100 frames sampled at 10 Hz from the videos was added to the training and the remaining 100 frames were added to the test set. The code was modified and all the IDIAP dataset images (4000 (authentic) +9950 (unauthentic) =13950) were provided for training the SVM. The dimension of the features was reduced using the PCA algorithm. The test set samples of subjects were projected using the coefficients and mean of the training samples obtained using the PCA algorithm after their features were extracted. It was observed that all the test set samples were accurately classified by the SVM.

5. CONCLUSIONS

Although tremendous advancements are being carried out to enhance the capability of deep networks and the success stories of deep-learned networks in various re-

cent applications are known, the work proposed in this article uses a simpler handcrafted-based approach for face-antispoofing using a lightweight machine learning classifier. The selected descriptors to extract significant features efficiently possess the capability to represent the face images more informatively and thus can be classified more accurately. These features exhibit dynamic biometrical traits and can be used for low dataset images and unbalanced dataset samples. Thus the proposed MLbFA model is more proficient which consolidates the advantages of handcrafted features and supervised learning with lower complexity. The result showed that the proposed MLbFA model achieved remarkable performance with a simpler feature extraction mechanism and classification.

The work will be extended for cross-dataset test samples. It can be tested for other types of attacks incorporating a fusion of conventional and blind features.

6. REFERENCES:

- [1] B. Chen, X. Qi, Y. Zhou, G. Yang, Y. Zheng, B. Xiao, "Image splicing localization using residual image and residual-based fully convolutional network", *Journal of Visual Communication and Image Representation*, Vol. 73, 2020, p. 102967.
- [2] Y. Tadmor, D. Tolhurst, "Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes", *Vision Research*, Vol. 40, No. 22, 2020, pp. 3145-3157.
- [3] Md R. Hasan, S. M. H. Mahmud, X. Y. Li., "Face Antispoofing using texture-based techniques and filtering methods", *Journal of Physics: Conference Series*, Vol. 1229, 2019.
- [4] C. Xin, W. Hongfei, Z. Jingmei, C. Hui, Z. Xiangmo, J. Yilin, "DTFA-Net: Dynamic and Texture Features Fusion Attention Network for Face Antispoofing", *Complexity*, Vol. 2020, 2020, p. 5836596.
- [5] P. Keyurkumar, H. Hu, J. A. Kumar, "Secure Face Unlock: Spoof Detection on Smartphones", *IEEE Transactions on Information Forensics and Security*, Vol. 11, No. 10, 2016, pp. 2268-2283.
- [6] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, R. Singh, "Computationally efficient face spoofing detection with motion magnification", *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Portland, OR, USA, 23-28 June 2013, pp. 105-110.
- [7] G. Pan, L. Sun, Z. Wu, S. Lao, "Eye blink-based anti-spoofing in face recognition from a generic web camera", *Proceedings of the 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, 14-21 October 2007, pp. 1-8.
- [8] I. Chingovska, A. Anjos, S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing", *Proceedings of the IEEE International Conference of Biometric Special Interest Group*, Darmstadt, Germany, 6-7 September 2012, pp. 1-7.
- [9] J. Maatta, A. Hadid, M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis", *Proceedings of the International Joint Conference on Biometrics*, Washington, DC, USA, 11-13 October 2011, pp. 1-7.
- [10] W. Bao, H. Li, N. Li, W. Jiang, "A liveness detection method for face recognition based on optical flow field", *Proceedings of the International Conference on Image Analysis and Signal Processing*, Linhai, China, 11-12 April 2009, pp. 233-236.
- [11] M. De Marsico, M. Nappi, D. Riccio, J. L. Dugelay, "Moving face spoofing detection via 3D projective invariants", *Proceedings of the International Conference on Biometrics*, New Delhi, India, 29 March - 1 April 2012, pp. 73-78.
- [12] D. Yuting, Q. Tong, X. Ming, Z. Ning, "Towards face presentation attack detection based on residual color texture representation", *Security and Communication Networks*, 2021, p. 6652727.
- [13] B. Zinelabidine, K. Jukka, H. Abdenour, "Face Anti-spoofing using Speeded-Up Robust Features and Fisher Vector Encoding", *IEEE Signal Processing Letters*, Vol. 24, No. 2, 2017, pp. 141-145.
- [14] F. Peng, L. Qin, M. Long, "CCoLBP: Chromatic Co-Occurrence of Local Binary Pattern for Face Presentation Attack Detection", *Proceedings of the 27th International Conference on Computer Communication and Networks*, Hangzhou, China, 30 July - 2 August 2018, pp. 1-9.
- [15] F. Peng, L. Qin, M. Long, "Face presentation attack detection based on chromatic co-occurrence of local binary pattern and ensemble learning", *Journal of Visual Communication and Image Recognition*, Vol. 66, 2020, p. 102746.
- [16] K. Mohan, P. Chandrashekhar, K. V. Ramanaiah, "Object-specific face authentication system for

- liveness detection using combined feature descriptors with fuzzy-based SVM classifier”, *International Journal of Computer Aided Engineering and Technology*, Vol. 12, No. 3, 2020, pp. 287-300.
- [17] A. Mittal, P. Kaur, A. Oberoi, “Hybrid Algorithm for Face Spoof Detection”, *International Journal for Research in Applied Science and Engineering Technology*, Vol. 10, No. 2, 2022.
- [18] M. Alshaiikhli, O. Elharrouss, S. Al-Maadeed, A. Bouridane, “Face-fake-net: the deep learning method for image face AS detection”, *Proceedings of the IEEE 9th European Workshop on Visual Information Processing*, Paris, France, 23-25 June 2021.
- [19] J. Zhou, K. Shu, P. Liu, J. Xiang, S. Xiong, “Face AS Based on Dynamic Color Texture Analysis Using Local Directional Number Pattern”, *Proceedings of the 25th International Conference on Pattern Recognition*, Milan, Italy, 10-15 January 2021, pp. 4221-4228.
- [20] C. Pei, Q. Hui-min, “Face AS algorithm combined with CNN and brightness equalization”, *Journal of Central South University*, Vol. 28, 2021, pp. 194-204.
- [21] Balamurali K, Chandru S, Muhammed Sohail Razvi and V. Sathiesh Kumar, “Face Spoof Detection Using VGG-Face Architecture”, *Journal of Physics: Conference Series*, Vol. 1917, 2021.
- [22] Z. Ming, M. Visani, M. M. Luqman, J.-C. Burie, “A Survey on AS Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices”, *Journal of Imaging*, Vol. 6, No. 12, 2020, p. 139.
- [23] Md R. Hasan, S. M. H. Mahmud, X. Y. Li, “Face AS Using Texture-Based Techniques and Filtering Methods”, *Journal of Physics: Conferences Series*, Vol. 1229, 2019.
- [24] P.T. de Freitas, K. Jukka, A. Andre, M. Jose De Mario, H. Abdenour, P. Matti, M. Sebastien, “Face liveness detection using dynamic texture”, *EURASIP Journal on Image and Video Processing*, Vol. 2, 2014.
- [25] A. Wetzler, R. Kimmel, “Efficient Beltrami Flow in Patch-Space. In *Scale Space and Variational Methods in Computer Vision. SSVN 2011*”, *Lecture Notes in Computer Science*, Vol. 6667, Springer, 2012, pp. 134-143.
- [26] M. Haghghat, S. Zonouz, M. Abdel-Mottaleb, “CloudID: Trustworthy cloud-based and cross-enterprise biometric identification”, *Expert Systems with Applications*, Vol. 42, No. 21, 2015, pp. 7905-7916.
- [27] A. Agarwal, R. Singh, M. Vatsa, A. Noore, “Boosting Face Presentation Attack Detection in Multi-Spectral Videos through Score Fusion of Wavelet Partition Images”, *Frontiers in Big Data*, Vol. 5, No. 22, 2022, p. 836749.
- [28] X. Tu, Z. Ma, J. Zhao, G. Du, M. Xie, J. Feng, “Learning generalizable and identity-discriminative representations for face anti-spoofing”, *ACM Transactions on Intelligent Systems and Technology*, Vol. 11, No. 5, 2020, pp. 1-19.
- [29] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, G. Zhao, “NASFAS: Static-dynamic central difference network search for face anti-spoofing”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 9, 2021, pp. 3005-3023.
- [30] T.-H. Lai, C.-Y. Peng, C.-L. Chou, “Fast Face Presentation Attack Detection in Thermal Infrared Images Based on Morphological Filtering”, *International Journal of Network Security*, Vol. 25, No. 2, 2023, pp. 185-193.
- [31] S. D. Thepade, M. R. Dindorkar, P. R. Chaudhari, S. V. Bang, “Enhanced Face Presentation Attack Prevention Employing Feature Fusion of Pretrained Deep CNN Model and Thepade’s Sorted Block Truncation Coding”, *International Journal of Engineering, Transactions A: Basics*, Vol. 36, No. 04, 2023, pp. 807-816.
- [32] A. H. Alharbi, S. Karthick, K. Venkatachalam, M. Abouhawwash, D. S. Khafaga, “Spoofing Face Detection Using Novel Edge-Net Autoencoder for Security”, *Intelligent Automation & Soft Computing*, Vol. 35, No. 3, 2023, pp. 2773-2787.
- [33] Z. Du, J. Li, L. Zuo, L. Zhu, K. Lu, “Energy-based domain generalization for face anti-spoofing”, *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, Portugal, 10-14 October 2022, pp. 1749-1757.
- [34] X. Chen, J. Zhou, X. Zhao, H. Wang, Y. Li, “A presentation attack detection network based on dynamic convolution and multilevel feature fusion with security and reliability”, *Future Generation Computer Systems*, Vol. 146, 2023, pp. 114-121.

- [35] Q. Zhou, K.-Y. Zhang, T. Yao, X. Lu, R. Yi, S. Ding, L. Ma, "Instance-aware domain generalization for face anti-spoofing", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17-24 June 2023, pp. 20453-20463.
- [36] H. Qi, R. Han, Y. Shi, X. Qi, "A Novel High-Performance Face Anti-Spoofing Detection Method", IEEE Access, Vol. 12, 2024, pp. 67379-67391.
- [37] J. Zhang et al. "UCDCN: a nested architecture based on central difference convolution for face-antispoofing", Complex and Intelligent Systems, Vol. 10, 2024, pp. 4817-4833.
- [38] M. Prasad, S. Jain, P. Bhanodia, A. Priya, "Influence of Standalone and Ensemble Classifiers in Face Spoofing Detection using LBP and CNN Models", European Journal of Electrical Engineering and Computer Science, Vol. 8, No. 2, 2024, pp. 17-30.

A Hybrid Deep Learning Framework for Speech-to-Text Conversion as Part of Telemedicine System Integrated With 5G

Review Paper

Medapati Venkata Manga Naga Sravan*

Andhra University
sravan.medapati@gmail.com

K Venkata Rao

HOD, Dept of Computer Science, Andhra University
professor_venkat@yahoo.com

*Corresponding author

Abstract – In today's world, aligning healthcare research with the third sustainable development goal of the United Nations (UN) is crucial. This goal focuses on ensuring health and well-being for all. Technological innovations like the Internet of Things (IoT) and Artificial Intelligence (AI) are vital in improving healthcare systems. Developing a technology-driven telemedicine system can have a significant impact on society. While current approaches focus on various methods for developing telemedicine modules, advancing these models with the latest technology is essential. Our paper proposes a deep learning-based framework that allows patients to provide information through voice. The system automatically analyzes this information to provide valuable insights in the doctor's dashboard, making diagnosis and prescriptions easier for the patient. Our proposed hybrid deep learning framework integrates with 5G technology and focuses on speech-to-text conversion. We introduce a hybrid deep learning model to improve performance in speech-to-text conversion. Our proposed algorithm, AI-Enabled Speech-to-Text Conversion (AIE-STTC), has the potential to match and surpass many existing deep learning models. Our empirical study, conducted using a benchmark dataset, demonstrated an impressive accuracy rate of 95.32%. In comparison, the baseline models showed lower accuracy rates: CNN achieved 88%, ResNet50 reached 90%, and VGG16 had 89%. Therefore, our proposed methodology has the potential to realize a technology-driven telemedicine system by integrating it with other necessary modules in the future. It significantly improves remote patient healthcare, making it more accessible and cost-effective, leading to a hopeful paradigm shift in healthcare services.

Keywords: Telemedicine System, Artificial Intelligence, 5G Technology, Deep Learning, Patient Voice-to-Text Conversion

Received: September 9, 2024; Received in revised form: December 23, 2024; Accepted: December 24, 2024

1. INTRODUCTION

Health and well-being for all are critical sustainable development goals the United Nations sets. Researchers have been developing various technologies and approaches to improve healthcare services in alignment with this goal. Traditional healthcare systems have been enhancing service delivery and disease diagnosis due to healthcare equipment and technologies innovations. However, the experience of the COVID-19 pandemic made the world more conscious about people's health regardless of their country or region. Efforts have been made to explore different means of providing healthcare services, including remote patient monitoring with the help of artificial intelligence

and Internet of Things technology, as well as creating a personalized medicine system by integrating required technologies such as 5G.

Regarding telemedicine, the Indian government has been working to provide healthcare services through public healthcare units accessible to people from all walks of life with just a phone call. However, there are many challenges in realizing such a telemedicine system. These challenges include technical issues related to infrastructure development, software integration, and security and privacy of existing healthcare-related IT systems. Regulatory challenges include licensing and credentialing, reimbursement policies, legal and compliance issues, and problems associated with user acceptance and experience, accessibility, provider train-

ing, and patient engagement. Clinical and operational challenges include technical support, secure data management, and quality of patient care. Furthermore, ethical challenges include patient privacy and ensuring equitable access. Economic factors associated with the telemedicine system, such as cost-benefit analysis and implementation costs, also pose challenges.

Many ongoing research efforts focus on providing advanced health services using 5G technology. Privacy concerns must be explored in a telemedicine system with 5G technology [1]. Integrating telemedicine with 5G and blockchain technology aims to enhance security, privacy, and non-repudiation [2]. Additionally, exploring a decision support system for telemedicine, incorporating edge computing and 5G technology along with sensors and available devices, is being pursued [3]. Technologies such as artificial intelligence and the Internet of Things are also being investigated as part of developing telemedicine systems to improve healthcare services [4]. Intelligent healthcare systems utilizing IoT technology and wearable devices are also being studied to integrate them with telemedicine and explore the ecosystem's role, including cloud, artificial intelligence, and artificial technologies [5]. Finally, a technical framework with various components is being explored to identify and plan the development of telemedicine systems for the future, leveraging innovative infrastructures, including 5G technology [6]. Based on the literature review, it is evident that developing a telemedicine system is a complex process, requiring investigation of various methods and gradual development to realize the potential of exploiting emerging technologies and advancing healthcare services.

This paper proposes a deep learning-based framework that enables patients to provide information through voice. The system automatically analyses this information to provide valuable insights on the doctor's dashboard, making diagnosis and prescriptions easier for the patient. Our hybrid deep learning framework integrates with 5G technology and focuses on speech-to-text conversion. We propose a hybrid deep learning model to enhance performance in speech-to-text conversion. Our algorithm, AI-Enabled Speech Conversion (AIESC), has the potential to outperform many existing deep learning models, as our empirical study using a benchmark dataset showed an impressive 95.32% accuracy rate. Therefore, our methodology has the potential to realize a technology-driven telemedicine system by integrating it with other necessary modules in the future. It significantly improves remote patient healthcare, making it more accessible and cost-effective, leading to a hopeful paradigm shift in healthcare services. The remaining sections of the paper are structured as follows: Section 2 reviews existing methods available in the literature for developing various modules in the telemedicine system. Section 3 presents the proposed deep learning-based or AI-enabled framework, which facilitates remote patients accessing

healthcare services. Section 4 presents the experimental results from our empirical study. Section 5 discusses the significance of the proposed system and its limitations. Section 6 concludes our work and provides directions for the future scope of the research.

2. RELATED WORK

Various existing methods are found in the literature about developing modules required for telemedicine systems. Lin *et al.* [1] improved connectivity between devices and energy efficiency with a 5G network. 5G enables safe, anonymous identity management for privacy in telemedicine, which improves the healthcare industry. Hameed *et al.* [2] proposed an IoHT-based healthcare system that combines blockchain, NN, and 5G for illness severity assessment and prediction. It improves healthcare efficiency and guarantees data confidentiality and privacy with 98.98% accuracy. PSO technology use, patient profiling, optimization, and various algorithms are examples of future developments. Wang *et al.* [3] suggested a 5G MEC-based telemedicine architecture incorporating OpenEMR with wearables. The multi-layered technique improves efficiency, scalability, and connection for applications other than Afib detection. Yu *et al.* [4] suggested a cloud-converged Internet of Things health architecture that prioritizes emotional engagement and multimodal sensing. A QoS framework for LAN-based health on the Internet of Things has been created. Suleiman *et al.* [5] examined how developments in 5G, IoT, AI, telemedicine, and networked competent healthcare are combined. It talks about difficulties, advantages, and potential futures.

Sadia *et al.* [6] included strategic planning, collaboration with medical professionals, and ongoing eHealth literacy. Public and private sector investments are necessary for affordable telemedicine in rural places with limited infrastructure. Research is essential for sustainable eHealth infrastructure, and 5G's role in connection is critical. Li *et al.* [7] assessed and validated the viability, effectiveness, and improved safety features of a 5G Telemedicine Network Latency Management System for telesurgery. Lin *et al.* [8] suggested a user-controlled single sign-on (SC-UCSSO) for telemedicine systems based on smartcards that ensure increased security, privacy, and performance. Hewa *et al.* [9] promoted using blockchain, 5G, and Multi-access Edge Computing (MEC) in digital healthcare infrastructure to improve patient privacy, data integrity, and scalability. Lu *et al.* [10] established the safety and efficacy of a telemedicine system for managing several diseases in a confined area.

Chettri *et al.* [11] concentrated on developing 5G wireless communication technologies that use Filter Bank Multicarrier (FBMC) for telemedicine effective transmission. The suggested method improves data rates to enable prompt patient monitoring. The simulation findings indicate possibilities for incorporating smartphones in telemedicine systems, increased effi-

ciency, and decreased delays. Diong *et al.* [12] exploited 5G to effectively handle the necessity for telemedicine in high-speed situations. By reducing needless handovers by 80.3%, a suggested changeover algorithm improves the quality of telemedicine services. Sadia *et al.* [13] presented a 5G healthcare system that is more efficient and less prone to latency than 4G using the TRILL protocol for data transport and mobility management. Figueiredo *et al.* [14] achieved fast speeds of 115 ps for an ultrafast electro-optical switch based on a chip-on-carrier semiconductor optical amplifier. Alenoghena *et al.* [15] examined eHealth, wireless technology, communication protocols, and problems in light of the COVID-19 pandemic's spike in telemedicine.

Lin *et al.* [16] presented an ID-based secure communication technique that protects privacy in 5G-IoT telemedicine systems. It integrates telemedicine with emergency medical services (EMS) and ensures the safe transfer of patient information, prompt delivery of emergency signals, and resilience to possible assaults. Liou *et al.* [17] suggested an affordable QoS benchmark system with good performance, simulating 5000 telemedicine devices for 5G uRLLC and mMTC scenarios. Adarsh *et al.* [18] suggested using effective communication technologies, dynamic prioritization, health service prioritization, and a cognitive radio-based telemedicine network for e-health. The performance of the proposed scenario can be improved by implementing WiMAX connectivity for mobility speeds less than 300 kmph and integrating 5G. The network may be further enhanced at the PCC level by utilizing SRD and UWB technology. Colella *et al.* [19] increased forecast precision, lowered expenses, and guaranteed successful BLM production quality. With its excellent parameter optimization performance, the suggested systematic design approach may be used for various industries and light sources. Silva *et al.* [20] presented a Local 5G Operator (L5GO) architecture emphasizing robotic surgery and augmented reality for delay-critical telemedicine. Regarding latency, the suggested L5GO performs better than conventional and Multi-access Edge Computing (MEC) networks, providing unique benefits for telehealth applications sensitive to delays.

Mihuba *et al.* [21] presented a mobile terminal and general packet radio service-based remote medical monitoring system that enables quick and affordable wireless telemedicine. The suggested architecture improves mobility and convenience by integrating sensors, CPUs, and communication. Bailo *et al.* [22] accessed healthcare expanded by telemedicine, which was essential during the epidemic. Lawmakers must support telesurgery since it presents both technological and legal issues. Chettri *et al.* [23] emphasized using Filter Bank Multicarrier (FBMC) in 5G telemedicine to improve remote healthcare in underprivileged regions by transmitting vital signs and imaging data efficiently. Arunsundar *et al.* [24] suggested integrating telemedicine into 5G networks to handle emergencies using

massive MIMO and cognitive radio networks. Peralta-Ochoa *et al.* [25] examined how 5G technology may be used in intelligent healthcare applications, focusing on theoretical ideas and small-scale applications. The research indicates that intelligent healthcare is becoming increasingly important, especially in light of the COVID-19 pandemic. A SWOT analysis to evaluate technical support and suggest alternatives may be part of future efforts.

Cabanillas-Carbonell *et al.* [26] examined 66 pertinent articles about how 5G could affect healthcare applications, focusing on cloud, AI, and IoT technology. The evaluation has significance for forthcoming investigations that seek to augment healthcare via 5G technologies, cultivating more intelligent, effective, and enduring healthcare systems. Albahri *et al.* [27] assessed networks, services, and applications related to IoT in telemedicine. By highlighting effective telemedicine for larger populations using IoT technology, it unearths answers from 141 publications. Ahmad *et al.* [28] Despite its importance during COVID-19, telehealth and telemedicine confront obstacles. Blockchain improves data security and privacy in healthcare by providing decentralized, traceable, and secure solutions. Jain *et al.* [29] proposed a 5G Network Slice-based digital system for real-time patient-centric healthcare to meet the post-COVID healthcare demand. Sadia *et al.* [30], Adford *et al.* [31] developed speech models trained on 680,000 hours of diverse data without fine-tuning. However, they acknowledged the need for extensive datasets and aimed to improve model accuracy and robustness in future research.

Based on the literature review, developing a telemedicine system is a complex process requiring investigation of various methods and gradual development to realize the potential of exploiting emerging technologies and advancing healthcare services. Orynbay *et al.* Specifically, [32] take a very newfound look to the integrations/synthesizes of speech, text and vision modalities, and mediates the devoted multi-modal interaction systems as a consequence. Abstract This paper is a review of recent advances in novel methods and technologies that facilitate joint, bidirectional communication between multiple modalities, through the enhancement of a single modality through behaviour image generation or through a more connected multi-sensory experience between modalities. In the study by Dhakad and Singh [33] the authors have provided a survey and performance analysis of speech to text technologies implemented by using python and their performance analysis, characteristics domain of usage. The paper evaluates different tools, frameworks focuses on their performances on Accuracy and Usability. Madhusudhana Reddy *et al.* Deep learning-based methods for speech-to-text and text-to-speech recognition are discussed to improve the performance and accuracy [34]. It identifies breakthrough neural models that are leading to the development of speech and text process-

ing systems. An extensive survey is provided by Sethiya and Maurya [35], which introduces advanced neural architectures for end-to-end speech-to-text translation systems. Authors define important challenges, methods, and future directions for work toward continuous, multilingual translation systems. Korchynskiy *et al.* [36] propose a method for enhancing the quality of speech-to-text conversion by noise suppression and language modelling. The true potential of the study is to understand it and make it easy, accurate, and energy-efficient for several applications. Telemedicine combines medical aid with technology during catastrophes, providing a tactical method for practical victim assessment, treatment prioritization, and coordination. Dar and Pusharaj [37] proposed a CNN-BLSTM hybrid model with Connectionist Temporal Classification (CTC) for speech recognition. The model achieved a word error rate of 36.97% but noted accuracy and training time challenges. Baevski *et al.* [38] introduced wav2vec 2.0, which performs well in voice recognition with less labelled data while identifying pre-training reliance as a limitation.

3. PROPOSED FRAMEWORK

The telemedicine system is a complex phenomenon that involves various components, protocols, networks, and communication methods. With the emergence of technologies like 5G and Artificial Intelligence, it has become possible to implement complex systems that were not feasible before. However, due to its complexity, we are focused on developing different methods to realize a technology-driven telemedicine system. In other words, we are addressing the challenge of developing various strategies that are part of a technology-driven telemedicine system. This section introduces the proposed methodology, algorithm, and hybrid deep learning model involved in the proposed system.

3.1. PROBLEM DEFINITION

The telemedicine system proposal includes a crucial method for translating spoken English into English text. This module is essential for developing the telemedicine system, as discussed in section 3.2. The challenging problem addressed in this work involves developing a hybrid deep learning model to convert English speech into English text as part of the telemedicine system. Additionally, our proposed telemedicine system will require other modules, the implementation of which will be deferred to our future work.

3.2. OUR FRAMEWORK

A technology-driven telemedicine system is envisioned as a game changer in health service provision for the general public. Many minor ailments may not require an in-person visit to a doctor. Providing affordable healthcare services to accommodate people from various economic backgrounds is crucial, reducing unnecessary expenses and time wastage. Commuting and spending a whole day to consult a doctor for a simple ailment that could be addressed through a phone consultation could be more efficient. Therefore, there is a need to develop a novel healthcare system, such as a telemedicine system, to enable people to seek advice from doctors without spending excessive money or time. Although developing a telemedicine system is complex, we propose a technology-driven one and implement one of the modules discussed in this paper. Implementing other models or methods necessary for actualizing a telemedicine system is deferred for our future endeavours. Fig. 1 displays the technology framework that leverages artificial intelligence and 5G technology to develop a telemedicine system.

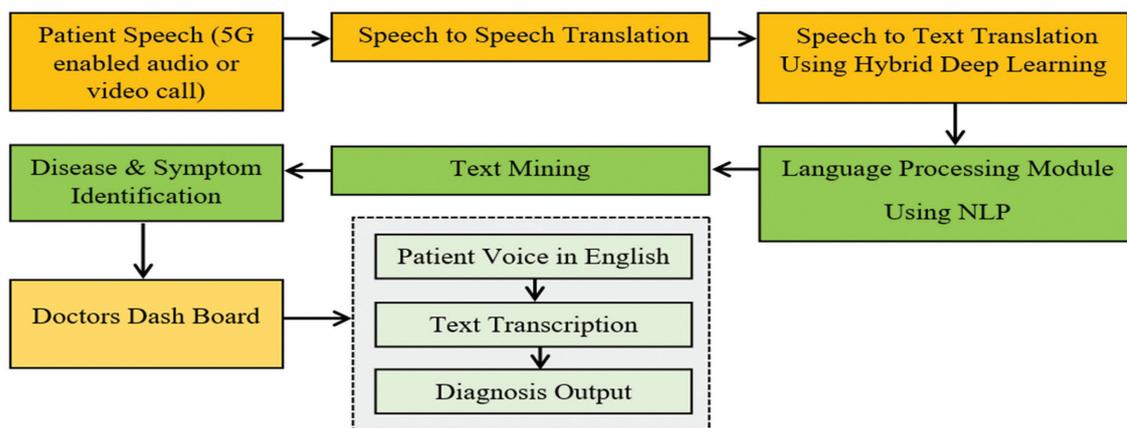


Fig. 1. Overview of the proposed telemedicine system

The proposed system aims to provide inclusive healthcare services accessible to patients of any region, religion, or language. The system is designed to understand spoken language, translate it, and provide information to doctors, making it easier for them to prescribe treatment. Patients can access healthcare

services through a simple phone call and receive necessary prescriptions or advice, significantly impacting the lives of people in society. The system utilizes voice-to-voice translation using deep learning models to translate a patient's native language into English accurately. Once translated, the speech is converted

to text, enabling advanced natural language processing and machine learning techniques for data analytics. The system also employs artificial intelligence to identify possible symptoms and diseases, presenting information to the doctor's dashboard. Without a doctor, healthcare professionals can handle patient calls, record the conversation, and make the data analytics results available to the doctor for decision-making. The doctor can then prescribe a solution or medicine, and the patient is informed immediately, either verbally or via messaging in their native language, eliminating the need to visit a healthcare facility physically.

The system ensures that patients can access healthcare services based on a set protocol, regardless of location. Integration with cloud technology allows doctors to access patient information whenever necessary. At the front end, the system receives patient calls in their native language and translates them into English, filtering background noise and using speech signals for accuracy. Deep learning models, including hybrid deep learning, recognize speech patterns and essential features for speech recognition. Language models play a significant role in understanding English text, preprocessing, and aiding machine learning models in data analytics. The proposed system uses artificial intelligence to streamline diagnosis, empowering doctors to make well-informed decisions.

3.3. 5G TECHNOLOGY

Technologies like 5G play a crucial role in developing telemedicine systems. This technology enables data transfer at a much faster rate compared to its predecessors. In other words, this technology allows patients to have video calls to consult with doctors, so the doctors can not only listen to the patients but also see the patient's condition better. The 5G technology enables high-quality data transfer, which is crucial for realizing a telemedicine system. When this technology is spread to

remote areas, it helps people communicate seamlessly with doctors through telemedicine. The access to the telemedicine system by people from all walks of life will be improved with 5G technology. Therefore, 5G technology can be adopted to improve healthcare services.

Considering the COVID-19 pandemic, where the world has learned a lesson about the importance of health, it is crucial to understand the importance of consulting doctors without physically moving to hospitals. Therefore, people from different fields need seamless access to the telemedicine system. People in remote areas can quickly access healthcare services through the telemedicine system. The 5G technology can also enable mobile health applications that provide interactivity between people and healthcare professionals. The technology can also be used to develop remote patient monitoring systems with the help of the Internet of Things and artificial intelligence to monitor patient vitals in real time and provide appropriate medical intervention. In the contemporary era, 5G technology is also being used for virtual surgeries from remote areas due to its real-time approach, low latency, and very high speed of data, making it possible to have remote-controlled systems with robotics for disease diagnosis and performing surgical procedures as well.

3.4. ENGLISH SPEECH TO ENGLISH TEXT TRANSLATION

We used an artificial intelligence-enabled approach with a hybrid deep learning model to convert patient speech into English text. The process, shown in Fig. 2, includes training and testing. In the first phase, the hybrid deep learning model is trained with features extracted from the training dataset, a speech recognition challenge dataset. After preprocessing the extracted features as spectrograms, they are used to train the model. The trained model can then translate any patient's speech into English text.

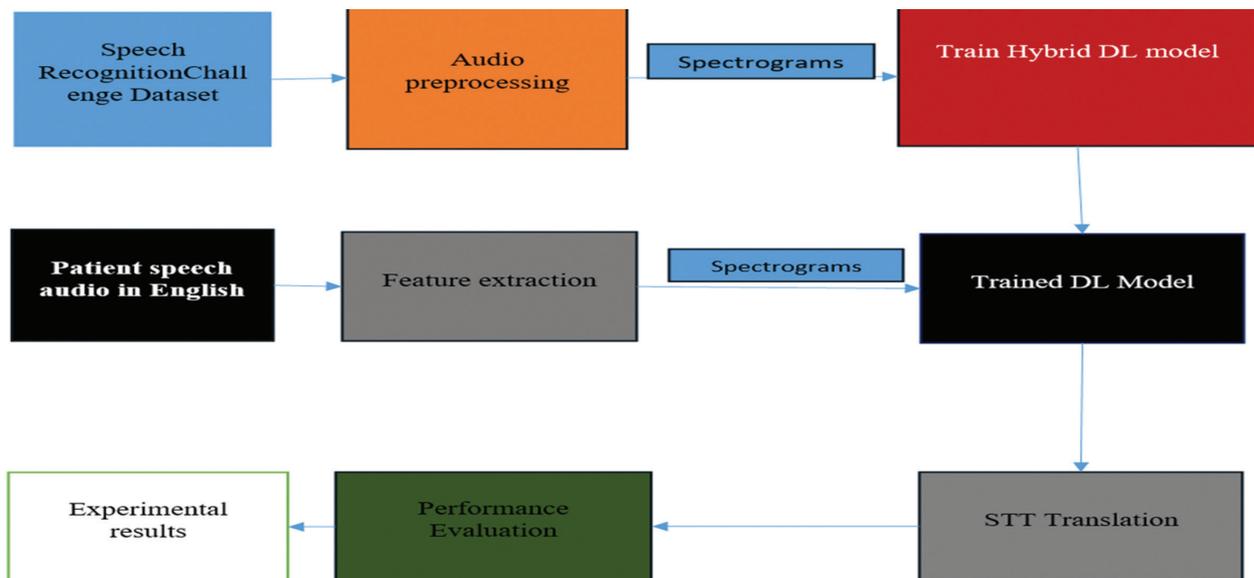


Fig. 2. Proposed methodology for speech-to-text (STT) conversion

Our previous discussion acknowledged that creating an entire telemedicine system is a complex task that demands significant resources.

Therefore, the development and empirical study outlined in this paper is focused solely on one module of the proposed telemedicine system: the conversion of English speech to English text. As previously mentioned, the patient communicates in their native language, which is then translated into English speech.

The development of this particular aspect of the framework is deferred to our future work. The primary focus of this paper is the translation of English speech to English text, which is accomplished through an enhanced deep learning model, as illustrated in Fig. 3. The deep learning model we have devised is a hybrid model that effectively utilizes convolutional layers and bidirectional GRU layers to efficiently convert English speech to text.

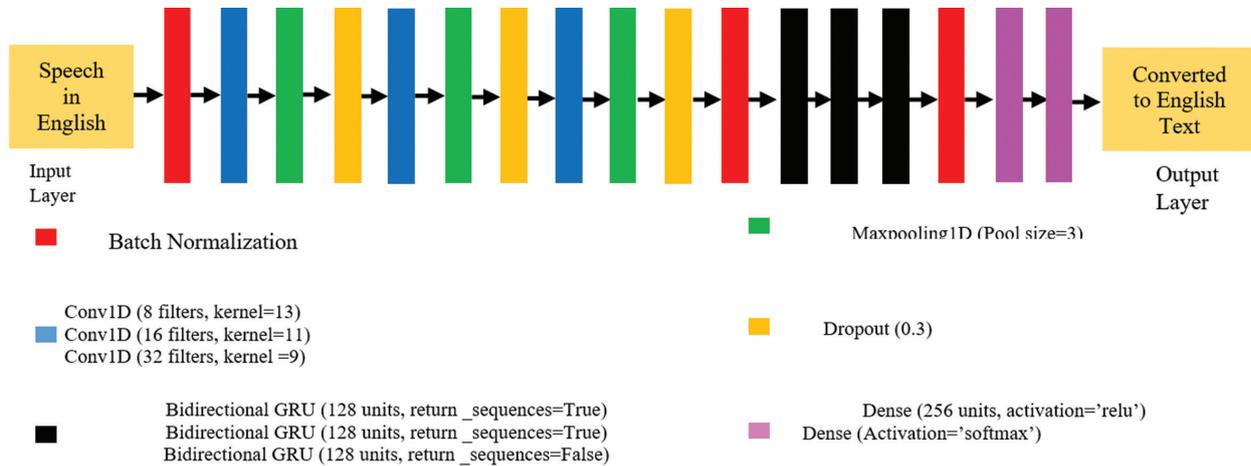


Fig. 3. The proposed hybrid deep learning model, a significant advancement for translating speech to English in a telemedicine system

In Fig. 3, we have a hybrid deep learning model designed for translating spoken English into written English text, specifically for use in a telemedicine system. The English speech input is fed into the model through the input layer, which then normalizes the input data for improved training performance and stability. The model employs convolutional layers, such as Conv1D with different filter sizes and kernels, and Maxpooling1D with a specified pool size to reduce the dimensionality of the feature maps. Dropout rate of 0.3 is used to address overfitting. The data is processed through several Bidirectional GRU layers with different configurations, including returning or not returning sequences. Finally, the model uses Dense layers with ReLU and softmax activations for the final output, which is the translated text in English. In summary, the model employs convolutional layers, max pooling, dropout, bidirectional GRUs, and dense layers to effectively translate spoken English into written English, enhancing the telemedicine system's ability to process and comprehend speech.

When combining CNN and bidirectional GRUs, batch normalization is essential for improving converting speech-to-text efficiency. The normalization process works on inputs of layers to leverage speed in training and achieve model stability. For speech recognition, normalization enhances the capability of the deep learning model. We are learning complex patterns from the data to make more accurate transcription. The normalization process is essential for leveraging the mod-

el's performance in the proposed hybrid architecture. In the proposed hybrid deep learning model, convolutional layers extract feature maps from the audio input. The filters used in the convolutional layers are meant to detect particular patterns in the voice and help improve transcription performance. The feature maps developed by convolutional layers help understand the phonetic elements in the given inputs and contextual information. Therefore, convolutional layers are crucial to understanding the difference between similar words and dealing with complexities in spoken language.

The proposed hybrid deep learning model also uses Max pooling layers, which take the feature maps obtained from convolutional layers and reduce spatial dimensions of the outcomes of convolutional layers. These layers use a sliding window on the given feature maps to get a value for which the max pooling is intended. This process involved in Max Pooling enables the model to reduce the feature maps and optimize for further processing. The method of lowering feature maps or optimizing them has its influence on reducing computational complexity, besides helping the hybrid model reduce overfitting problems, as it provides a concept known as translation invariants that helps understand unseen data. As deep learning models are extended neural networks, they are designed to eliminate overfitting problems. This reduction of overfitting is achieved with dropout layers, which can help improve the learning process and reduce noise, besides addressing the issue of overfitting by setting some

inputs to zero in the training process towards making the model more robust in learning and understanding from the data.

A Bidirectional Gated Recurrent Unit (BiGRU) is used in the proposed hybrid deep learning model to enhance the capability of the model to understand dependencies in the sequential nature of data. The BiGRU has two layers that work in the forward and backward directions. This dual approach can help understand the context of the past and the future states while dealing with language data, which is essentially time series data. The proposed deep learning model has a fully connected layer, an essential component in the network. Every neuron is connected to every other neuron in the preceding layer to ensure holistic interaction and data flow among the layers. This layer helps understand the complex patterns related to voice data towards converting data from speech to English text.

3.5. PROPOSED ALGORITHM

The proposed algorithm, AI-Enabled Speech-to-Text Conversion (AIE-STTC), is a critical component of this research. It aims to develop a mechanism to automatically convert patient speech into text using a hybrid deep learning model. The algorithm takes the patient's speech in audio format and the training data to train the deep learning model. Once trained, the model can automatically convert the patient's speech into English text. This algorithm is designed to be a part of a Telemedicine system, offering numerous benefits for healthcare services.

Algorithm: AI-Enabled Speech to Text Conversion (AIE-STTC)

Input: Patient speech audio q , training dataset T

Output: STT conversion results in R , performance statistics P

1. Begin
2. Initialize features map M
3. For each sample t in T
4. $features \leftarrow \text{ExtractFeatures}(t)$
5. Add t and features to M
6. End For
7. Configure hybrid DL model m (as in Fig. 3)
8. Compile m
9. $m' \leftarrow \text{TrainDLModel}(m, M)$
10. Persist m'
11. Load m'
12. $R \leftarrow \text{STTConversion}(m', q)$
13. $P \leftarrow \text{Evaluation}(R, \text{ground truth})$
14. Print R
15. Print P
16. End

Algorithm 1: AI-Enabled Speech-to-Text Conversion (AIE-STTC)

Algorithm 1 is designed for speech-to-text conversion, an essential component of the telemedicine system. It employs an AI-enabled approach using a hybrid deep learning model illustrated in Figure 3. The algorithm involves training the deep learning model with a provided training dataset. Before teaching the model, there is a preprocessing step where each training sample is converted into spectrograms. These features are then used to train the proposed deep learning model. The training process involves using all the samples in the training dataset, where each sample consists of the patient's speech and the corresponding converted English text.

This helps the deep learning model to learn from the samples and acquire sufficient knowledge. Once the model has gained knowledge, it is saved for future use. When the algorithm receives a new patient's speech as input, it is converted into English text using the trained deep learning model. The algorithm progresses toward achieving patient speech-to-text conversion, ultimately benefitting further modules associated with that element. During testing, the deep learning model demonstrates its capability to translate patient speech into English text, and the algorithm assesses the model's performance. This algorithm enables the realization of the speech-to-text conversion module in the telemedicine system, which plays a crucial role in the data analytics module. Implementing various modules within the telemedicine system can help improve the quality of healthcare services. This novel, technology-driven approach can bring significant benefits to people at large.

3.6. DATASET DETAILS

This paper's empirical study uses the benchmark data set collector from [39].

4. EXPERIMENTAL RESULTS

This section presents the results of our empirical study. The proposed telemedicine system consists of multiple modules, with the STT conversion module being the focus of this paper's empirical study. This module enables the telemedicine system to translate the patient's speech audio into English transcription. To train the deep learning model, we utilized a dataset comprising pairs of samples, each containing speech audio and its corresponding text. The training set shall comprise 80% of the data, while the testing set comprises 20%. The number of epochs used for model training is 10, while the learning rate is 0.001. Given that the input is audio content, the proposed algorithm leverages MFCC features to train the model. It is then saved and reused for new patients to convert their speech audio into English transcription. The English text obtained is subsequently utilized by other modules, such as the data analytics module, to aid in identifying patient diseases and symptoms. This supports doctors in the process of diagnosis and prescription.

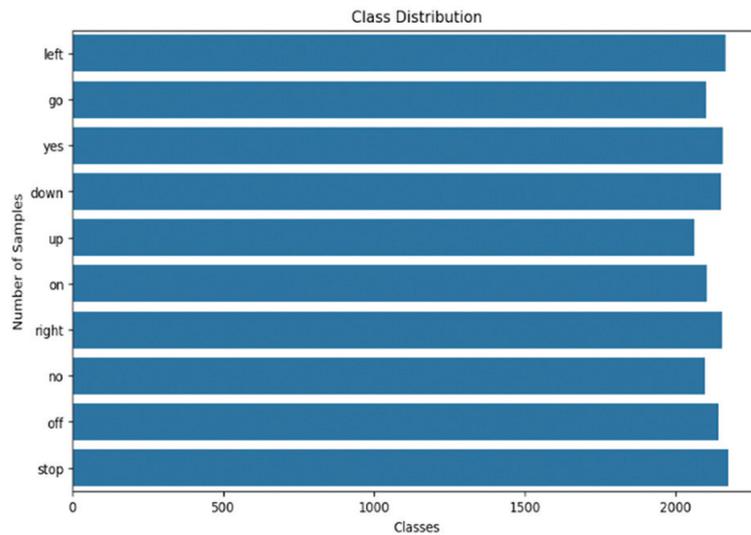


Fig. 4. Class distribution dynamics of the dataset

Fig. 4 illustrates the distribution of classes in the dataset using a bar graph. The horizontal axis represents the class counts, while the vertical axis represents the class samples. Each class in the dataset contains a tentative

range of 1800 to 2200 samples. This balanced dataset provides valuable support for artificial intelligence models to learn from the data and effectively perform their tasks.

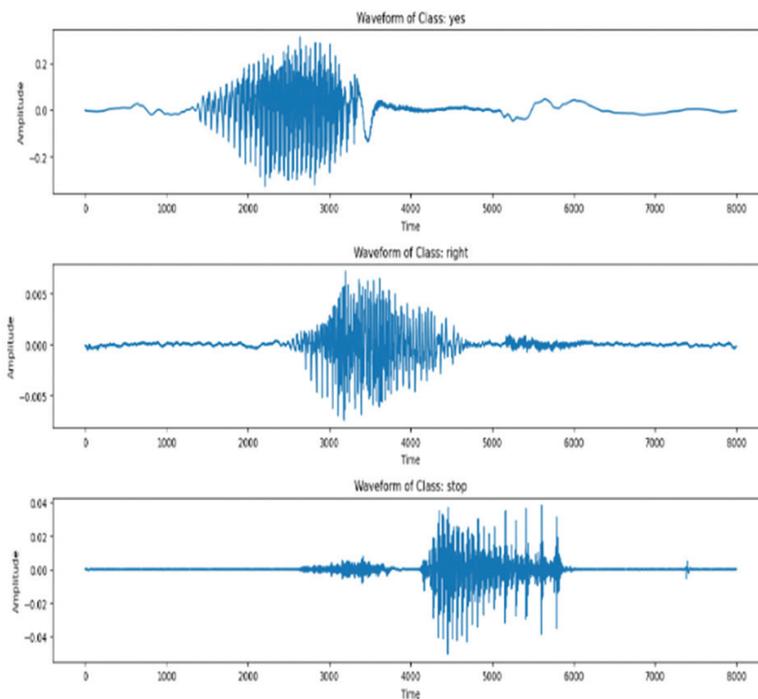


Fig. 5. Waveforms of some audio samples

In Fig. 5, waveforms of various media samples illustrate the amplitude of signals over time. Each sample shows a different amplitude waveform. These waveforms offer essential clues for deep learning models to understand language better and convert it to English text. Additionally, the waveforms provide temporal dynamics related to the audio content, giving valuable insights to artificial intelligence models that aim to capture the essence of audio samples during training and speech-to-text conversion processes.

Fig. 6 shows spectrograms of given audio samples. The proposed telemedicine system uses a hybrid deep learning model for speech-to-text conversion. Before training the model, features are extracted from the audio using spectrograms, visual representations associated with a spectrum of frequencies linked to a given audio signal. Spectrograms aid in understanding time-frequency analysis, enabling the model to convert speech audio into English text. An audio signal is a continuous waveform reflected in sound pressure vari-

ations. Short-Time Fourier Transform (STFT) is the underlying technique used to create spectrograms from given audio samples, facilitating further processing by

deep learning models. Fig. 7 shows various distribution dynamics and also skewness distribution dynamics.

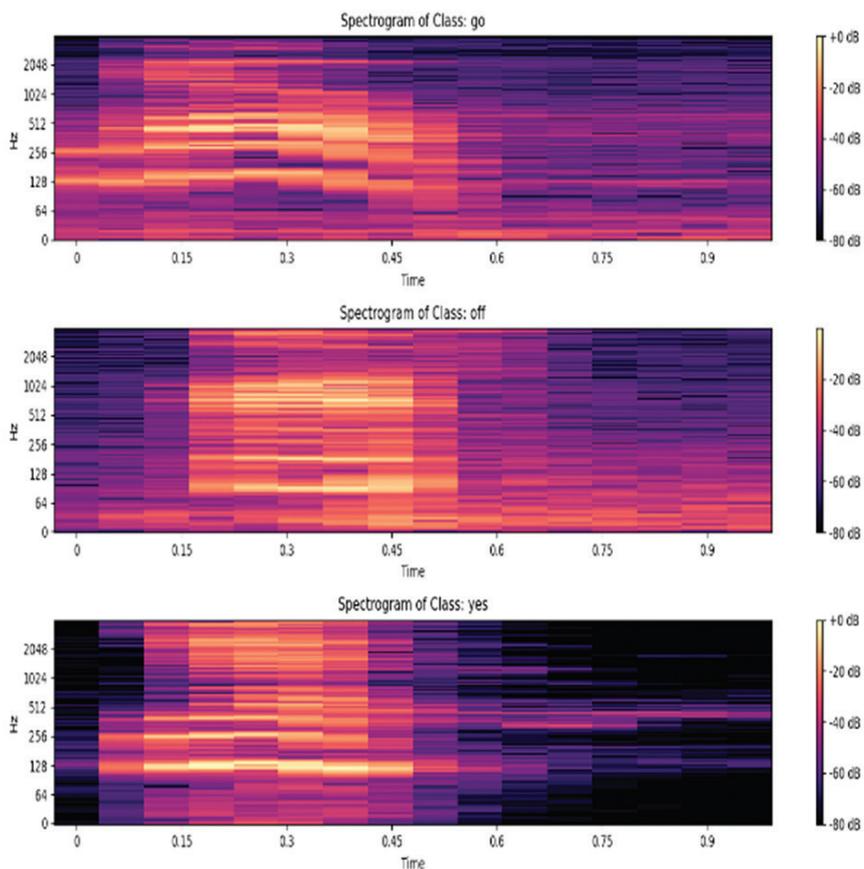


Fig. 6. Spectrogram of audio samples

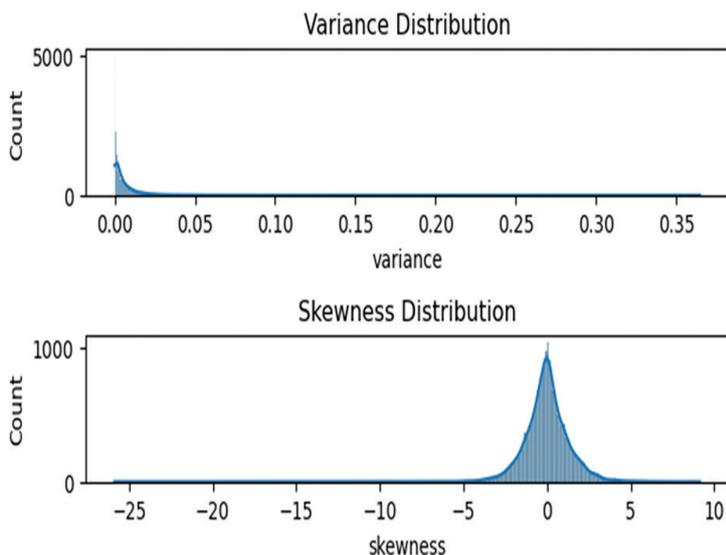


Fig. 7. Variance and skewness distribution dynamics

The variance distribution of a given audio sample visualizes how the features appear across various time windows in the frequency domain, aiding in understanding the differences in emotional tones or phonetic changes. The skewness distribution, visualized in the figure, measures asymmetry in the audio data.

A positive skewness indicates a longer tail on the right side, while a negative skewness indicates the opposite. Fig. 8 shows the proposed hybrid depth learning model's loss dynamics for STT conversion. The model's loss dynamics are shown against the number of epochs. Model loss visualization is provided for training and test

data over multiple learning cycles. The results indicate that the training performance remains consistent as the number of learning cycles increases. This means that the training loss gradually decreases until convergence as the number of epochs increases. For the test data, there are some fluctuations evident in the loss function, unlike the training data. However, there is an overall decrease in the loss as the number of epochs increases, indicating improvement in the model. It's important to note that while overall improvement in the model, there is some instability in the model's performance concerning the test data. Fig. 9 shows the model accuracy dynamics against several epochs.

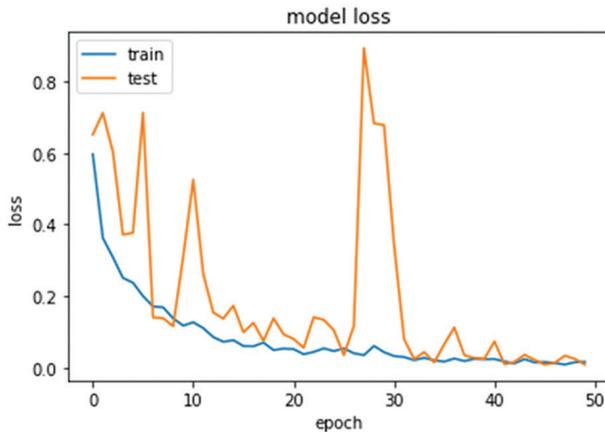


Fig. 8. The loss dynamics of the proposed deep learning model against the number of epochs

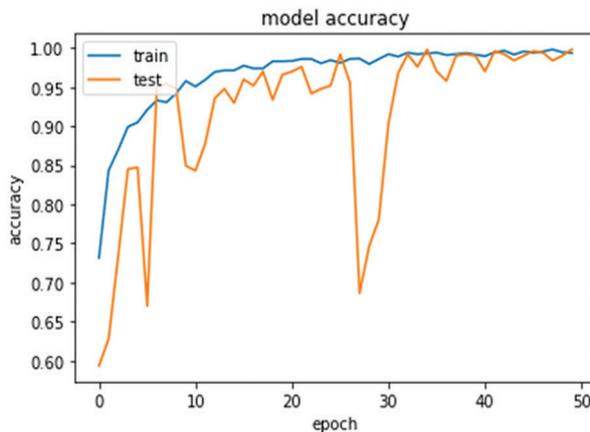


Fig. 9. Accuracy of the proposed hybrid deep learning model against several epochs

The proposed deep learning model's training and test accuracy are provided, showing that as the number of epochs increases, the accuracy of the training and test data gradually increases. The model demonstrated consistency in training accuracy, while there were fluctuations in test accuracy. Overall, the observations indicate that the proposed hybrid deep learning model consistently improves performance in terms of accuracy as the number of epochs increases until convergence. Fig. 10 compares deep learning models' performance in the STT conversion process.

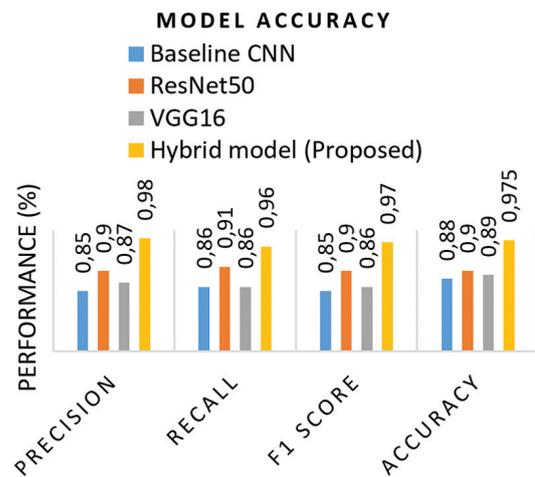


Fig. 10. Performance comparison among deep learning models in STT conversion

The hybrid deep learning model's performance is compared to state-of-the-art models. The existing deep learning models include baseline CNN and pre-trained models such as ResNet-50 and VGG-16. When all the models are used to convert the patient's speech into English, their performance is observed to be different due to their other methods of operation and as they have different layers in the deep learning process. The results show that the baseline CNN model achieved 85% precision, ResNet-50 90%, VGG-16 87%, and the proposed hybrid deep learning model achieved 98% precision. Regarding the recall measure, the baseline CNN model achieved 86%, ResNet-50 91%, VGG-16 86%, and the proposed deep learning model achieved 96% recall. Regarding the F1 score measure, the baseline CNN model achieved 85%, ResNet-50 90%, and VGG-16 86%, while the proposed deep learning model achieved a 97% F1 score. In terms of accuracy measure, the baseline CNN model achieved 88%, ResNet-50 90%, VGG-16 89%, and the proposed hybrid deep learning model achieved 97.50% accuracy. The results show that the proposed deep learning model achieved the highest accuracy, outperforming all the state-of-the-art models with 97.50%. Performance is also evaluated using the metric in Eq. 1.

$$WER = \frac{\text{Substitutions} + \text{Insertions} + \text{Deletions}}{\text{Total Words}} \quad (1)$$

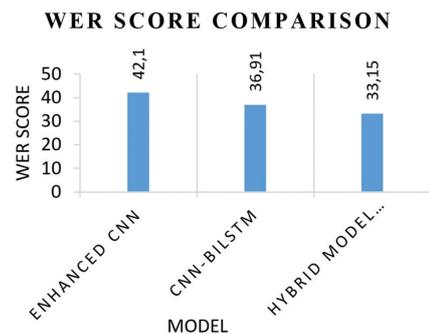


Fig. 11. WER score comparison

As presented in Fig. 11, the proposed model performs better than existing models, with a WER score of 33.15%, compared with the baseline CNN (42.1%) and CNN-BILSTM (36.91%) [38] model.

5. DISCUSSION

With the emergence of innovative technologies, telemedicine in various real-world applications is becoming increasingly important, especially in light of the recent lessons learned from the COVID-19 pandemic. The global understanding of the need for advanced healthcare systems to assist patients without requiring them to visit a hospital has become evident. Therefore, it is crucial to develop telemedicine systems that can save patients time, effort, and money for various reasons. Technological advancements such as the Internet of Things, Artificial Intelligence, and 5G technology have made exploring new avenues in remote patient monitoring and telemedicine systems possible, bringing healthcare services within easy reach with just a simple phone call. 5G technology enables seamless and efficient real-time communication between patients and doctors through wearable and non-wearable devices, including video calls. This allows doctors to listen to the patient's speech and see the patient live, enabling them to provide timely prescriptions. Patients can also have video calls, allowing the doctor to observe the patient for a more accurate diagnosis and treatment. This paper outlines a technology-driven architecture for a telemedicine system with multiple modules. Given the complexity of telemedicine systems, this paper focuses on one module - converting patient speech audio into English text, which other modules will use for data analytics to provide possible disease information and symptoms to the doctor. This will assist the doctor in taking the necessary steps for prescription and guiding the patient in overcoming their ailments. This telemedicine system can help remote patients access healthcare services without disrupting their daily activities, saving them significant time, effort, and money. While this paper proposes a technologically advanced telemedicine system, it is essential to note that it must be fully implemented. Only the speech-to-text conversion module has been implemented, while the implementation of other modules is deferred to future work. The implemented speech-to-text conversion module has been evaluated, and certain limitations have been identified, as discussed in section 5.1.

5.1. LIMITATIONS

The speech-to-text conversion module described in this paper is based on a hybrid deep learning model. The model has been evaluated, and its results are compared with state-of-the-art models. While the proposed model demonstrates superior accuracy compared to existing models, the speech conversion module has some limitations. A significant limitation is that it has been evaluated with limited samples in the

dataset. The findings can only be generalized with a diverse range of real-time patient speech samples. Another significant limitation is that the system has yet to be integrated with any existing healthcare applications used by healthcare units.

6. CONCLUSION AND FUTURE WORK

Our paper presents a framework based on deep learning that enables patients to provide information through voice. The system automatically analyzes this information and provides valuable insights on the doctor's dashboard, making diagnosis and prescriptions easier for the patient. Our proposed hybrid deep learning framework integrates with 5G technology and emphasizes speech-to-text conversion. We introduce a hybrid deep learning model to enhance performance in speech-to-text conversion. We propose an algorithm called AI-Enabled Speech Conversion (AIESC), which utilizes the improved hybrid deep learning model to convert speech to text efficiently. Using a benchmark dataset, our empirical study demonstrated that our proposed model outperforms many existing deep learning models with a 97.50% accuracy rate. In the future, we intend to improve the system by developing a method for converting patient speech to English speech, analyzing patient speech for disease diagnosis, and identifying various symptoms based on the patient's voice information. It is also desirable to integrate multiple methods involved in the telemedicine system to realize a complete and technology-driven telemedicine system that can serve remote patients without the need to visit healthcare facilities and incur significant expenses. This paradigm shift in healthcare services will be possible with an efficient telemedicine system integrated with 5G technology.

DECLARATION

FUNDING

No financial support was received by the authors in this research.

COMPETING INTERESTS

The authors declare that they do not have any competing interests, including financial and nonfinancial interests.

ETHICAL APPROVAL

This research does not involve humans or animals, so no ethical approval is required.

CONSENT FOR PUBLICATION

The authors give consent for their publication.

DATA AVAILABILITY

Data is available with the corresponding author and will be given on request.

AUTHOR CONTRIBUTION

All authors contributed to the study's conception and design. Medapati Venkata Manga Naga Sravan and Prof K Venkata Rao performed material preparation, data collection, and analysis. Medapati Venkata Manga Naga Sravan wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

7. REFERENCES

- [1] T.W. Lin, C. L. Hsu, "FAIDM for Medical Privacy Protection in 5G Telemedicine Systems", *Applied Sciences*, Vol. 11, No. 3, 2021, p. 1155.
- [2] K. Hameed, I. S. Bajwa, N. Sarwar, W. Anwar, Z. Mushtaq, T. Rashid, "Integration of 5G and Blockchain Technologies in Smart Telemedicine Using IoT", *Journal of Healthcare Engineering*, Vol. 2021, 2021, pp. 1-18.
- [3] Y. Wang, P. Tran, J. Wojtusiak, "From Wearable Device to OpenEMR: 5G Edge Centered Telemedicine and Decision Support System", *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies*, Vol. 5, 2022, pp. 491-498.
- [4] H. Yu, Z. Zhou, "Optimization of IoT-Based Artificial Intelligence Assisted Telemedicine Health Analysis System", *IEEE Access*, Vol. 9, 2021, pp. 85034-85048.
- [5] T. A. Suleiman, A. Adinoyi, "Telemedicine and Smart Healthcare - The Role of Artificial Intelligence, 5G, Cloud Services, and Other Enabling Technologies", *International Journal of Communication, Network and System Sciences*, Vol. 1, 2023, pp. 31-51.
- [6] A. Sadia, P. Ramjee, "Framework for Future Telemedicine Planning and Infrastructure using 5G Technology", *Wireless Personal Communications*, Vol. 100, 2018, pp. 193-208.
- [7] C. Li et al. "Telemedicine network latency management system in 5G telesurgery: a feasibility and effectiveness study", *Surgical Endoscopy*, Vol. 38, 2023, pp. 1592-1599.
- [8] T. W. Lin, C. L. Hsu, T. V. Le, C. F. Lu, B. Y. Huang, "A Smartcard-Based User-Controlled Single Sign-On for Privacy Preservation in 5G-IoT Telemedicine Systems", *Sensors*, Vol. 21, No. 8, 2021, p. 2880.
- [9] T. Hewa, A. Braeken, M. Ylianttila, M. Liyanage, "Multi-Access Edge Computing and Blockchain-based Secure Telehealth System Connected with 5G and IoT", *Proceedings of GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, 7-11 December 2020.
- [10] J. Lu, X. Wang, X. Zeng, W. Zhong, W. Han, "Application of telemedicine system on the management of general patient in quarantine", *Scientific Reports*, Vol. 13, 2023, p. 12215.
- [11] L. Chettri, "Design and Implementation of 5G Wireless Telemedicine Systems", *International Journal of Information Studies*, Vol. 8, No. 2, 2016, pp. 45-53.
- [12] B. W. Diong, M. I. Goh, S. K. Chung, A. Chekima, H. T. Yew, "Vertical Handover Algorithm for Telemedicine Application in 5G Heterogeneous Wireless Networks", *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 8, 2021, pp. 611-617.
- [13] D. Sadia, P. Anand, A. Awais, R. Seungmin, "Emerging Mobile Communication Technologies for Healthcare System in 5G Network", *Proceedings of the IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, Auckland, New Zealand, 8-12 August 2016, pp. 47-54.
- [14] R. C. Figueiredo, N. S. Ribeiro, A. M. O. Ribeiro, Cri, "5G-Enabled Health Systems: Solutions, Challenges and Future Research Trends", *Journal of Lightwave Technology*, Vol. 33, No. 1, 2019, pp. 1-9.
- [15] C. O. Alenoghena, H. O. Ohize, A. O. Adejo, "Telemedicine A Survey of Telecommunication Technologies, Developments, and Challenges", *Journal of Sensor and Actuator Networks*, Vol. 12, No. 2, 2023, p. 20.
- [16] T. W. Lin, "A Privacy-Preserved ID-Based Secure Communication Scheme in 5G-IoT Telemedicine Systems", *Sensors*, Vol. 22, No. 18, 2022, p. 6838.

- [17] E. C. Liou, S. C. Cheng, "A QoS Benchmark System for Telemedicine Communication Over 5G uRLLC and mMTC Scenarios", *Proceedings of the IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability*, Tainan, Taiwan, 29-31 May 2020, pp. 24-26.
- [18] A. Adarsh, S. Pathak, B. Kumar, "Design and Analysis of a Reliable, Prioritized and Cognitive Radio-Controlled Telemedicine Network Architecture for Internet of Healthcare Things", *International Journal of Computer Networks and Applications*, Vol. 8, No. 1, 2021, pp. 54-66.
- [19] R. Colella, L. Catarinucci, "10.525 GHz Backscattering RFID System Based on Doppler Radar Technology for 5G Applications and Telemedicine", *Proceedings of the Photonics & Electromagnetics Research Symposium*, Rome, Italy, 17-20 June 2019, pp. 3689-3695.
- [20] R. D. Silva, Y. Siriwardhana, T. Samarasinghe, M. Ylianttila, M. Liyanage, "Local 5G Operator Architecture for Delay Critical Telehealth Applications", *Proceedings of the IEEE 3rd 5G World Forum*, Bangalore, India, 10-12 September 2020, pp. 257-262.
- [21] G. G. Mihuba, M. J. Shundi, O. K. Bishoge, An, "Design of Telemedicine System Based on Mobile Terminal", *International Journal of Emerging Technologies in Engineering Research*, Vol. 7, No. 1, 2019, pp. 1-8.
- [22] P. Bailo, F. Gibelli, A. Blandino, A. Piccinini, "Telemedicine Applications in the Era of COVID-19 Telesurgery Issues", *International Journal of Environmental Research and Public Health*, Vol. 19, No.1, 2022, p. 323.
- [23] L. Chettri, R. Bera, "Design and Analysis of 5G Telemedicine Systems", *Mobile and Forensics*, Vol. 3, No. 2, 2021, pp. 48-57.
- [24] B. Arunsundar, R. Srinivasan, "Implementation of Obiquitous Telemedicine on 5g Networks", *International Journal of Advanced Research Trends in Engineering and Technology*, Vol. 2, No. 1, 2015, pp. 1-5.
- [25] A. M. Peralta-Ochoa, P. A. Chaca-Asmal, L. F. Guerrero-Vásque, "Smart Healthcare Applications over 5G Networks: A Systematic Review", *Applied Sciences*, Vol. 13, No. 3, 2023, p. 1469.
- [26] M. Cabanillas-Carbonell, J. Pérez-Martínez, J. A. Yáñez, "5G Technology in the Digital Transformation of Healthcare: A Systematic Review", *Sustainability*, Vol. 15, No. 4, 2023, p. 31278.
- [27] A. S. Albahri, J. K. Alwan, Z. K. Taha, S. F. Ismail, R. A. Hamid, A. A. Zaidan, O. S. Albahri, B. B. Zaidan, A. H. Alamoodi, M. A. Alsalem, "IoT-based telemedicine for disease prevention and health promotion: State-of-the-Art", *Journal of Network and Computer Applications*, Vol. 173, 2021, pp. 1-59.
- [28] R. W. Ahmad, K. Salah, R. Jayaraman, I. Yaqoob, S. Ellahham, M. Omar, "The role of blockchain technology in telehealth and telemedicine", *International Journal of Medical Informatics*, Vol. 148, 2021.
- [29] H. Jain, V. Chamola, Y. Jain, Naren, "5G network slice for digital real-time healthcare system powered by network data analytics", *Internet of Things and Cyber-Physical Systems*, Vol. 1, 2021, pp. 14-21.
- [30] A. Sadia, P. Ramjee, B. S. Chowdhary, M. R. Anjum, "A Telemedicine Platform for Disaster Management and Emergency Care", *Wireless Personal Communications*, Vol. 106, No. 1, 2019, pp. 191-204.
- [31] A. Adford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision", *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, HI, USA, 23-29 July 2023, pp. 28492-28518.
- [32] L. Orynbay, B. Razakhova, P. Peer, B. Meden, E. Žiga, "Recent Advances in Synthesis and Interaction of Speech, Text, and Vision. Electronics", *Electronics*, Vol. 13, No. 9, 2024, p. 1726.
- [33] A. Dhakad, S. Singh, "Python-Powered Speech-to-Text: A Comprehensive Survey and Performance Analysis", *International Journal of Engineering Research & Technology*, Vol. 12, No. 9, 2023.
- [34] V. M. Reddy, T. Vaishnavi, K. P. Kumar, "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning", *Proceedings of the 2nd International Conference on Edge Computing and Applications*, Namakkal, India, 19-21 July 2023.

- [35] N. Sethiya, C. K. Maurya, "End-to-End Speech-to-Text Translation: A Survey", arXiv:2312.01053, 2024.
- [36] V. Korchynskiy, I. Vynogradov, "Methods of Improving the Quality of Speech-to-Text Conversion", Scientific Collection InterConf, Vol. 194, 2024, pp. 426-437.
- [37] M. Ahmad Dar, J. Pushparaj, "Hybrid Architecture CNN-BLSTM for Automatic Speech Recognition", Proceedings of the 3rd International Conference on Artificial Intelligence for Internet of Things, Vellore, India, 3-4 May 2024, pp. 1-15.
- [38] A. Baeovski, Y. Zhou, A. Mohamed, M. Auli, "Wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations", Proceedings of Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 12449-12460.
- [39] TensorFlow Speech Recognition Challenge dataset, <https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/data> (accessed: 2024)

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia.

About this Journal

The International Journal of Electrical and Computer Engineering Systems publishes original research in the form of full papers, case studies, reviews and surveys. It covers theory and application of electrical and computer engineering, synergy of computer systems and computational methods with electrical and electronic systems, as well as interdisciplinary research.

Topics of interest include, but are not limited to:

- Power systems
- Renewable electricity production
- Power electronics
- Electrical drives
- Industrial electronics
- Communication systems
- Advanced modulation techniques
- RFID devices and systems
- Signal and data processing
- Image processing
- Multimedia systems
- Microelectronics
- Instrumentation and measurement
- Control systems
- Robotics
- Modeling and simulation
- Modern computer architectures
- Computer networks
- Embedded systems
- High-performance computing
- Parallel and distributed computer systems
- Human-computer systems
- Intelligent systems
- Multi-agent and holonic systems
- Real-time systems
- Software engineering
- Internet and web applications and systems
- Applications of computer systems in engineering and related disciplines
- Mathematical models of engineering systems
- Engineering management
- Engineering education

Paper Submission

Authors are invited to submit original, unpublished research papers that are not being considered by another journal or any other publisher. Manuscripts must be submitted in doc, docx, rtf or pdf format, and limited to 30 one-column double-spaced pages. All figures and tables must be cited and placed in the body of the paper. Provide contact information of all authors and designate the corresponding author who should submit the manuscript to <https://ijeces.ferit.hr>. The corresponding author is responsible for ensuring that the article's publication has been approved by all coauthors and by the institutions of the authors if required. All enquiries concerning the publication of accepted papers should be sent to ijeces@ferit.hr.

The following information should be included in the submission:

- paper title;
- full name of each author;
- full institutional mailing addresses;
- e-mail addresses of each author;
- abstract (should be self-contained and not exceed 150 words). Introduction should have no subheadings;
- manuscript should contain one to five alphabetically ordered keywords;
- all abbreviations used in the manuscript should be explained by first appearance;
- all acknowledgments should be included at the end of the paper;
- authors are responsible for ensuring that the information in each reference is complete and accurate. All references must be numbered consecutively and citations of references in text should be identified using numbers in square brackets. All references should be cited within the text;
- each figure should be integrated in the text and cited in a consecutive order. Upon acceptance of the paper, each figure should be of high quality in one of the following formats: EPS, WMF, BMP and TIFF;
- corrected proofs must be returned to the publisher within 7 days of receipt.

Peer Review

All manuscripts are subject to peer review and must meet academic standards. Submissions will be first considered by an editor-

in-chief and if not rejected right away, then they will be reviewed by anonymous reviewers. The submitting author will be asked to provide the names of 5 proposed reviewers including their e-mail addresses. The proposed reviewers should be in the research field of the manuscript. They should not be affiliated to the same institution of the manuscript author(s) and should not have had any collaboration with any of the authors during the last 3 years.

Author Benefits

The corresponding author will be provided with a .pdf file of the article or alternatively one hardcopy of the journal free of charge.

Units of Measurement

Units of measurement should be presented simply and concisely using System International (SI) units.

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003

Published: semiannually

Copyright

Authors of the International Journal of Electrical and Computer Engineering Systems must transfer copyright to the publisher in written form.

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.

Postal Address

Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b
31000 Osijek, Croatia

IJECES Copyright Transfer Form

(Please, read this carefully)

This form is intended for all accepted material submitted to the IJECES journal and must accompany any such material before publication.

TITLE OF ARTICLE (hereinafter referred to as "the Work"):

COMPLETE LIST OF AUTHORS:

The undersigned hereby assigns to the IJECES all rights under copyright that may exist in and to the above Work, and any revised or expanded works submitted to the IJECES by the undersigned based on the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the complete Work and all incorporated parts of the Work. Otherwise he/she warrants that necessary permissions have been obtained for those parts of works originating from other authors or publishers.

Authors retain all proprietary rights in any process or procedure described in the Work. Authors may reproduce or authorize others to reproduce the Work or derivative works for the author's personal use or for company use, provided that the source and the IJECES copyright notice are indicated, the copies are not used in any way that implies IJECES endorsement of a product or service of any author, and the copies themselves are not offered for sale. In the case of a Work performed under a special government contract or grant, the IJECES recognizes that the government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official government purposes only, if the contract/grant so requires. For all uses not covered previously, authors must ask for permission from the IJECES to reproduce or authorize the reproduction of the Work or material extracted from the Work. Although authors are permitted to re-use all or portions of the Work in other works, this excludes granting third-party requests for reprinting, republishing, or other types of re-use. The IJECES must handle all such third-party requests. The IJECES distributes its publication by various means and media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various collections, databases and other publications. The IJECES publisher requires that the consent of the first-named author be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes. If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IJECES publisher assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

Authors of IJECES journal articles and other material must ensure that their Work meets originality, authorship, author responsibilities and author misconduct requirements. It is the responsibility of the authors, not the IJECES publisher, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

- The undersigned represents that he/she has the authority to make and execute this assignment.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
- The undersigned agrees to indemnify and hold harmless the IJECES publisher from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

Author/Authorized Agent

Date

CONTACT

International Journal of Electrical and Computer Engineering Systems (IJECES)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek
Kneza Trpimira 2b
31000 Osijek, Croatia
Phone: +38531224600,
Fax: +38531224605,
e-mail: ijeces@ferit.hr