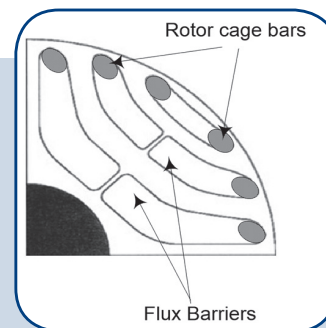
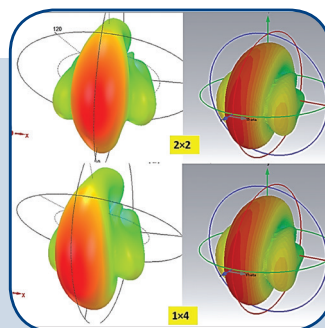
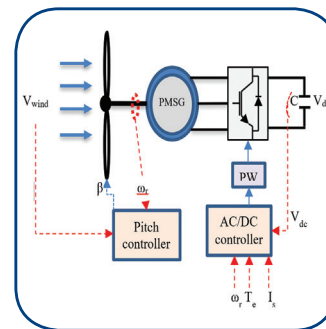
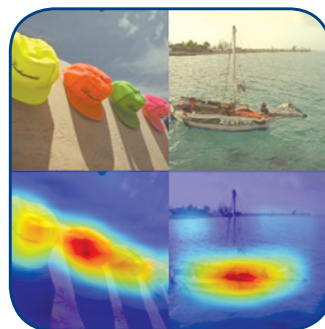
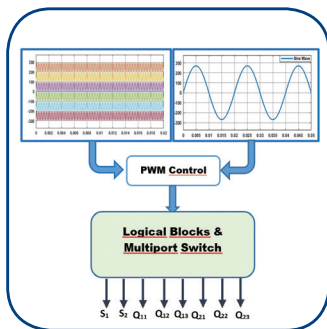


International Journal of Electrical and Computer Engineering Systems



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 14, Number 5, 2023 | Pages 497 - 612

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

MANAGING EDITOR

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

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

EDITORIAL BOARD

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

Leo Budin
University of Zagreb, Croatia

Matjaz Colnarič
University of Maribor, Slovenia

Aura Conci
Fluminense Federal University, Brazil

Bojan Čukić
West Virginia University, USA

Radu Dobrin
Malardalen University, Sweden

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

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

Ratko Grbić
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

Dražan Kozak
J.J. Strossmayer University of Osijek, Croatia

Sven Lončarić
University of Zagreb, Croatia

Tomislav Kilić
University of Split, Croatia

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

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

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

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

Davor Pavuna

Ecole Polytechnique Fédérale de
Lausanne, Switzerland

Nedjeljko Perić
University of Zagreb, Croatia

Marjan Popov
Delft University, The Netherlands

Sasikumar Punnekkat
Mälardalen University, Sweden

Chiara Ravasio
University of Bergamo, Italy

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

Gregor Rozinaj
Slovak University of Technology, Slovakia

Imre Rudas
Budapest Tech, Hungary

Ivan Samardžić
J.J. Strossmayer University of Osijek, Croatia

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

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

Cristina Seceleanu
Mälardalen University, Sweden

Siniša Srblić
University of Zagreb, Croatia

Zdenko Šimić
University of Zagreb, Croatia

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

Domen Verber
University of Maribor, Slovenia

Dean Vučinić
Vrije Universiteit Brussel, Belgium
J.J. Strossmayer University of Osijek, Croatia

Joachim Weickert
Saarland University, Germany

Drago Žagar
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

Design and Simulation of Microstrip Antenna Array Operating at S-band for Wireless Communication System	495
<i>Original Scientific Paper</i> Sattar Othman Hasan Saman Khabbat Ezzulddin Rashad Hassan Mahmud Mowfaq Jalil Ahmed	
Noise Effects on a Proposed Algorithm for Signal Reconstruction and Bandwidth Optimization	507
<i>Original Scientific Paper</i> Ahmed F. Ashour Ashraf A. M. Khalaf Aziza I. Hussein Hesham F. A. Hamed Ashraf Ramadan	
Lucy Richardson and Mean Modified Wiener Filter for Construction of Super-Resolution Image	517
<i>Original Scientific Paper</i> Pravin Balaso Chopade Prabhakar N. Kota Bhagvat D. Jadhav Pravin Marotrao Ghate Shankar Dattatray Chavan	
JPEG2000-Based Semantic Image Compression using CNN	527
<i>Original Scientific Paper</i> Anish Nagarsenker Prasad Khandekar Minal Deshmukh	
Enhancing Dynamic Hand Gesture Recognition using Feature Concatenation via Multi-Input Hybrid Model	535
<i>Original Scientific Paper</i> Djazila Souhila Kortti Zohra Slimane Kheira Lakhdari	
Performance Analysis of a new Filter and Wrapper Sequence for the Survivability Prediction of Breast Cancer Patients	547
<i>Original Scientific Paper</i> E. Jenifer Sweetlin S. Saudia	
A New Approach using Deep Learning and Reinforcement Learning in HealthCare: Skin Cancer Classification	557
<i>Original Scientific Paper</i> Dahdouh Yousra Anouar Boudhir Abdelhakim Ben Ahmed Mohamed	
Real-World Anomaly Detection in Video Using Spatio-Temporal Features Analysis for Weakly Labelled Data with Auto Label Generation	565
<i>Original Scientific Paper</i> Rikin J. Nayak Jitendra P. Chaudhari	
Development of a Control Strategy for the Hybrid Energy Storage Systems in Standalone Microgrid	575
<i>Original Scientific Paper</i> Hocine Guentri Abdeldjalil Dahbi Tayeb Allaoui Salim Aoulmit Ahmed Bouraiou	
Performance of Synchronous Reluctance Generators with Series and Shunt Stator Connections	585
<i>Case Studie</i> Pauline Ijeoma Obe Lilian Livutse Amuhaya Emeka Simon Obe Adamu Murtala Zungeru	
Design and analysis of a new multi-level inverter topology with a reduced number of switches and controlled by PDPWM technique	593
<i>Case Studie</i> Fatima Chakir Abdelmounime EL Magri Rachid Lajouad Mohamed Kissaoui Mostafa Chakir Omar Bouattane	
Biomolecular Event Extraction using Natural Language Processing	601
<i>Review Paper</i> Manish Bali S.P. Anandaraj	
About this Journal	
IJECES Copyright Transfer Form	

Design and Simulation of Microstrip Antenna Array Operating at S-band for Wireless Communication System

Original Scientific Paper

Sattar Othman Hasan

Salahaddin University-Erbil,
College of Education, Department of Physics
Erbil, Iraq
sattar.hasan@su.edu.krd

Saman Khabbat Ezzulddin

Salahaddin University-Erbil,
College of Science, Department of Physics
Erbil, Iraq

Rashad Hassan Mahmud

Salahaddin University-Erbil,
College of Education, Department of Physics
Erbil, Iraq

Mowfaq Jalil Ahmed

Salahaddin University-Erbil,
College of Education, Department of Physics
Erbil, Iraq

Abstract – In this article, different design configurations of rectangular microstrip patch antenna (RMSA) array operating at S-band frequency are presented. The substrate material utilized in the designs is Rogers-RT-5800 with dielectric permittivity ($\epsilon_r=2.2$), thickness of ($h=1.6$ mm), and loss tangent of ($\delta = 0.009$). The performances of a single element, (1×2), (2×2) and (1×4) array elements operating at (3.6 GHz) are investigated using the CST and HFSS numerical techniques. The simulation results indicates that the antenna gain of (8.68, 10.35, 10.43 and 10.52) dB, VSWR (1.045, 1.325, 1.095 and 1.945), return loss (-34.91, -17.15, -27.42 and -12.26) dB, and bandwidth (85.00, 200.00, 215 and 106.4) MHz are achieved with the implementation of HFSS for advanced single element, (1×2), (2×2) and (1×4) array elements, respectively. Besides, the corresponding antenna parameter values provided by CST are, gain (7.36, 9.8, 9.87 and 10.30) dB, VSWR (1.011, 1.304, 1.305 and 1.579), return loss (-44.97, -17.58, -17.55 and -14.01) dB, and bandwidth (92.28, 204, 229.49 and 129.12) MHz, respectively. The results also reveals that the higher gain and wider bandwidth are, respectively, achieved with (1×4) and (2×2) array configuration arrangement and with both simulation techniques. Additionally, a good agreement and an advancement between the obtained results with the ones previously studied for the same array types operating at S-band frequencies are also observed.

Keywords: Patch Antenna, Microstrip antenna, antenna array, wireless communication, s-band.

1. INTRODUCTION

In the field of wireless communications, the antennas have played a significant role, and have many forms including patch antennas, wire antennas, horn antennas and parabolic antennas, having own properties and applications and without which the world could never reached at this stage of developed technology systems [1,2]. In the last two decades, the wireless communication systems have been developed from analog into digital technologies systems to improve the data rate capability and increase the speed of multimedia transmission [3]. New 5G frequency bands which utilized (3.3-3.8) GHz in several countries around the world, for the 6 GHz sub-band, have been approved to support higher data speed transfer [3,4].

On the other hand, as technological innovations for instance smart phones, Internet-of-Things (IoT), wear-

able and handheld devices become smaller over the time, it is critical yet challenging to develop an antenna geometrical arrangement that maintains at a particular band frequency providing high directivity, gain, wider bandwidth, and high-quality services [1,5]. Therefore, the MSA can be regarded as a reliable candidate for the new wireless application systems due to its low cost, light weight and ease of fabrications. However, the MSA has a number of drawbacks, including low gain and narrow bandwidth which make the researcher to develop methods to overcome these MSA disadvantages.

Many studies were published in the literature proposing different feeding methodologies, slots, defected ground plane, various antenna arrangement arrays to achieve antenna improvement characteristics [6,7]. Among these methods, the MSA arrays of various geometrical construction have been increasingly

employed in recent years, since they are durable and provide very attractive performance for several application systems such as, 5G, satellite, medical applications, WLAN, radar, personal technology and military application systems [8-10]. A planar slot RMSA array operating at (3.5 GHz) was designed and fabricated by [11] for a sub-6 GHz 5G wireless application systems. It was found that the proposed antenna provides a gain, efficiency and bandwidth of the order of (4.2 dB), (82%) and (19 MHz), respectively. A RMSPA gain improvement through an array of (4x1) configuration operating at (2.5 GHz) was designed and simulated on the FR4 dielectric substrate of thickness (1.6 mm) making it reliable for WiMAX applications [7]. Besides, a higher radiation performance for an inset fed single element, (1x2) and (1x4) arrays RMSPA placed on Rogers-RT-5880 with thickness (1.6 mm) operating at (2.4 GHz) for WLAN applications was achieved by [12] using IE3D numerical method. Additionally, a single RMSA array with (1x2) and (2x2) arrays was proposed to operate at S-band frequencies using FR4 dielectric substrate of thickness (1.6 mm) [13,14]. On the other hand, a coaxial probe fed single RMSPA, (1x4), (2x2) and (4x4) element arrays were designed and investigated by [15] using RT-Duriod dielectric substrates with thickness ($h=3.175$ mm) to operate at (1.48 GHz) for radar application systems. Moreover, the directivity enhancement of a triangular MSA operating at (5.5 GHz) was proposed and fabricated through a T-junction inset fed of (2x2) patch element arrays using FR4 epoxy as a substrate material with thickness (1.6 mm) [16]. Besides, the size reduction of RMSA array operating at (2.4 GHz) with acceptable radiation performance were investigated by [33] using different array configuration types. Generally, these research works displayed that the gain and bandwidth could be enhanced properly with the use of reliable array constructions.

Therefore, this work is established to improve the overall radiation performance of RMSA array operating at (3.6 GHz) through an investigation of various geometrical array configurations. The design procedure is performed by arranging four element RMSA array in a form of single patch, (1x2), (2x2), and (1x4) array geometrical configurations and the fundamental antenna parameters for each of them are calculated using both HFSS and CST simulation methods. The main purpose of the study is intended to attain a lightweight antenna providing reliable gain and wider bandwidth suitable for most S-band application systems which is a part of the electromagnetic spectrum ranged between (2-4) GHz. This band frequency is selected due to the fact that it implemented for radio and television, satellite communication, radar weather station, wireless network and ship radar [17,18].

The remind section of the article is organized as follows. In section 2, a theoretical design calculation of the physical dimension for single element, (1x2), (2x2), and (1x4) array array RMSA array elements are described. The computation of the fundamental RMSA

array parameters for each mentioned array configuration are presented and discussed in section 3. The main conclusion and recommendation for future work are mentioned in section 4.

2. MATERIAL METHOD

In this study, a single (MSA) is designed and then upgraded to the two and four (MSA) elements in both the parallel (corporate) with series feeding method, to improve the value of gain, directivity and bandwidth. The design procedure for all mentioned array cases is performed by the implementation of both CST and HFSS software package. These two simulation methods are, respectively, based on the numerical methods known as Finite Element Method (FEM) and Finite Integral Techniques (FIT) [19]. The (FEM) is an alternative computational electromagnetic method (CEM) suggested to determine a solution of partial differential equation in the frequency domain. It was proposed in the 1940s by Courant and it was firstly employed to solve problems containing difference in electric potential but it is now widely used in the field of RF/microwave system technology. In this technique, the total domain is discretized into a sub-domain of unordered mesh which is a successful solver for complicated structures and non-uniform objects which offers a high degree of geometric tolerance [20]. Meanwhile, in (FIT), the solution of Maxwell equations can be derived from both frequency and time domains and it was firstly approved by Weiland 1977 [21]. This method is suitable to different electromagnetic problems extending from static field measurements to high frequency applications in time as well as frequency domain. The (FIT) is composed of distinct types of grids, like non-orthogonal or Cartesian and describes a volume within Maxwell's equations which can be resolved in a finite unit of discrete locations. Hence, (FIT) is extremely equivalent to the finite difference time domain method (FDTD) [2].

Furthermore, both CST and HFSS are a full-wave electromagnetic field (EM) simulation software that helps to solve all 3D electromagnetic or mechanical second differential problems [2]. As the antenna operational frequency (f_r), dielectric substrate permittivity (ϵ_r) and substrate thickness (h) are specified, then the dimensions of patch and ground plane are computed through some mathematical expression provided by transmission line method. Later, the radiation performance for each of the mentioned array structures are computed using CST and HFSS simulation methods. The simulation design for single element RMSA and considered array configurations are described in the following subsections.

2.1. SINGLE ELEMENT

The design of single element RMSA with specification of its patch and fed line dimension is demonstrated in Figure (1) which are computed by implementing both mentioned simulation software techniques.

The proposed antenna is built up on Rogers-RT-5800 dielectric substrate of relative permittivity ($\epsilon_r=2.2$) and thickness of ($h=1.6$ mm) and having loss tangent of ($\delta=0.009$) [22,23]. In the design procedure, an extensively inset fed and transmission line feeding methods have been utilized and their patch and line dimensions are evaluated through the following equations based on transmission line method [11]. The first quantity is the patch width (W_p) and it can be determined using equation [24]:

$$W_p = \frac{c}{2f_r} \sqrt{\frac{2}{\epsilon_r+1}} \quad (1)$$

where, (f_r) is the antenna operating frequency, (c) is speed of light in free space and (ϵ_r) is the substrate relative permittivity. The second parameter that must be specified is the effective dielectric constant (ϵ_{eff}) which is introduced due to the fringing field effect and for RMSA is calculated as given by [24]:

$$\epsilon_{eff} = \frac{\epsilon_r+1}{2} + \frac{\epsilon_r-1}{2} \left[1 + 12 \frac{h}{W_p} \right] \quad (2)$$

As a consequence of radiation fringing effect on both sides of the patch, then the length is extended by an amount (ΔL) and is determined through an equation given by [24].

$$\Delta L = \frac{0.412(\epsilon_{eff}+0.3)\left(\frac{W_p}{h}+0.264\right)}{(\epsilon_{eff}-0.258)\left(\frac{W_p}{h}+0.8\right)} \epsilon_{eff} \quad (3)$$

and,

$$L_{eff} = \frac{c}{2f_r \sqrt{\epsilon_{eff}}} \quad (4)$$

where, (L_{eff}) is the patch effective length is expressed as:

While the actual patch length (L_p) can be obtained with respect to the extension and effective length of the patch as:

$$L_p = L_{eff} - 2\Delta L \quad (5)$$

To improve the performance of the patch antenna design, the ground plane should be larger than the patch dimensions by approximately six times as the substrate, hence the width (W_g) and length (L_g) of substrate are calculated by using the following relations [13,25]:

$$W_g = W_p + 2 \left(\frac{\lambda}{4} \right) \quad (6)$$

$$L_g = L_p + 2 \left(\frac{\lambda}{4} \right) \quad (7)$$

Moreover, the width of feed line (W_f) can be obtained from input impedance equations as mathematically expressed by [9] as:

$$Z_o = \begin{cases} \frac{60}{\sqrt{\epsilon_{reff}}} \ln \left[\frac{8h}{W_f} + \frac{W_f}{4h} \right] & \text{if } \frac{W_f}{h} \leq 1 \\ \frac{120\pi}{\sqrt{\epsilon_{reff}} \left[\frac{W_f}{h} + 1.393 + 0.667 \times \ln \left(\frac{W_f}{h} + 1.444 \right) \right]} & \text{if } \frac{W_f}{h} > 1 \end{cases} \quad (8)$$

where, (Z_o) is the antenna impedance, and for transmission line feed is equal to 50 ohms, while, the characteristics impedance (Z_o) of the two microstrip lines is given by [6] as:

$$Z_o = 2 \times 50 \Omega \quad (9)$$

Moreover, the length of the inset-fed line is specified by [7]:

$$F_i = \frac{6h}{2} \quad (10)$$

Depended on the above equations, the single RMSA dimensions after optimization through the mentioned simulation methods are determined and the results are summarized in the Table 1.

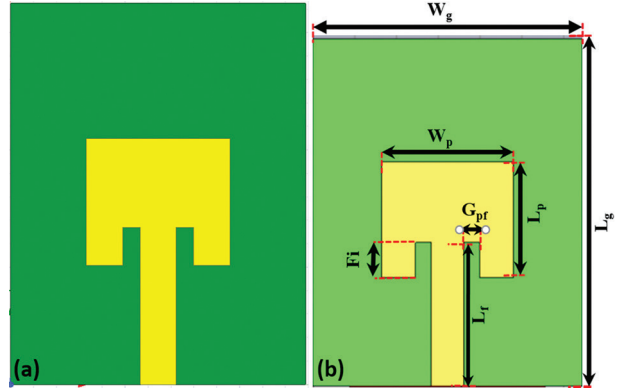


Fig. 1. Top view of simulated single RMSPA with (a) HFSS and (b) CST.

Table 1. Dimension optimization of single element

Parameters	Description	Value(mm)
W_p	Patch width	27.00
L_p	Patch length	31.48
G_{pff}	Gap between Line feed and Patch	3.75
F_i	Inset-feed length	8.15
L_f	Transition line length	33.06
W_f	Microstrip line feed width	6.75
W_g	Ground width	64.25
L_g	Ground length	82.06
H	Substrate thickness	1.575

2.2. ANTENNA ARRAY

As mentioned before, the antenna array is usually employed to enhance the antenna directivity, gain and plays different roles that are not achieved normally by a single element [26]. Generally, the array factor is depended on the array elements, space between element, geometry of the array and excitation vector. For one-dimensional linear array, the mathematical expression for the array factor is formulated as given by [27]:

$$AF(\theta, \phi) = \sum_{m=1}^M I_m e^{j(k \cdot \hat{r}_m + \alpha_m)} \quad (11)$$

While, for two-dimensional planner array, this factor is expressed as [27]:

$$AF(\theta, \phi) = \sum_{m=1}^M \sum_{n=1}^N I_{mn} e^{j(k \cdot \hat{r}_{mn} + \alpha_{mn})} \quad (12)$$

Where, (\hat{r}) is the elements vector location, (I_{mn}) is amplitude and (α_{mn}) is the excitation phase for (m^{th}, n^{th}) elements to achieve maximum directivity in the given direction (θ, ϕ) [28]. Regarding the parameters that have been evaluated for single patch element, a (1x2), (2x2) and (1x4) RMSA arrays are built. The rectangular patch elements are implemented in a linear and planer configuration, using inset line feed techniques with input impedance of (50Ω) . Each patch elements arrays are separated by (0.5λ) with optimized dimensions as demonstrated in Figures (2 and 3) which are designed with CST and HFSS. This arrangement has the benefit of becoming easier to establish as well as being efficiently optimized by the patch's inset-fed and length of the inset fed [29,30].

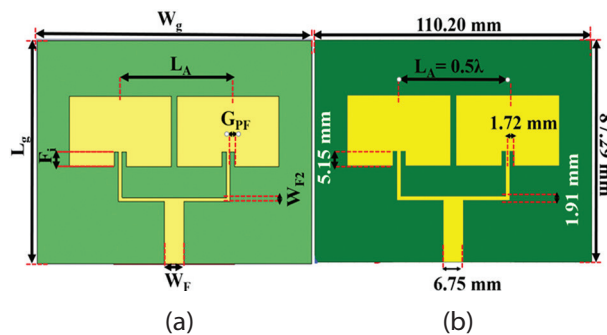


Fig. 2. Top view of proposed (1x2) RMSA arrays simulated by (a) CST and (b) HFSS.

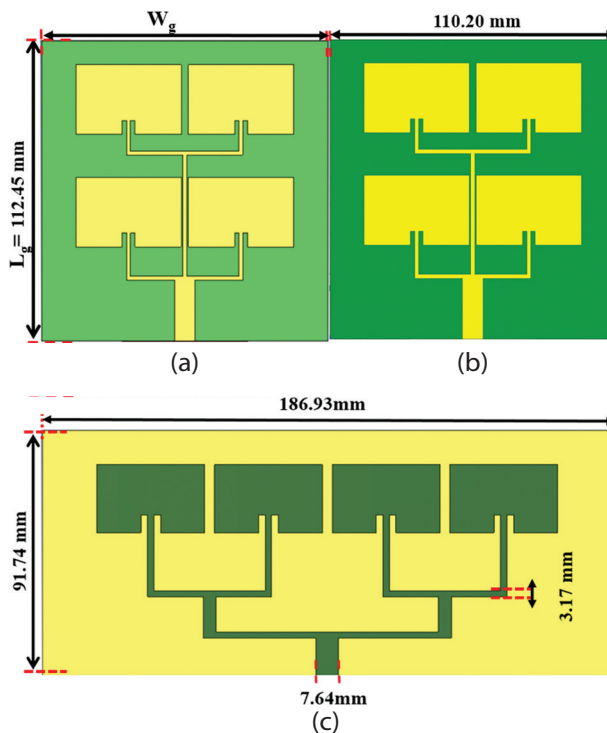


Fig. 3. Top view of proposed (2x2) RMSA arrays simulated by: (a) CST, (b) HFSS and (c) (1x4) by CST.

When the dimensions of the patch, ground plane, fed-line as well as dielectric substrate characteristics are specified, then the radiation performance of each array configuration are evaluated by both simulation

method and the result are presented in the form of table or graphs as described in the next section.

3. RESULTS AND DISCUSSION

As previously mentioned, the main purpose of the present work is to develop a high gain and wider bandwidth (RMSA) with lightweight as well as offering better radiation characteristics suitable for compact S-band wireless applications systems operating at (3.6) GHz. According to the optimized patch, ground plane and inset-fed dimension values of the proposed antennas, the reflection coefficient (S11), voltage standing wave ratio (VSWR), gain, directivity, bandwidth and antenna radiation pattern for each mentioned array configuration have been analyzed by both HFSS and CST computational methods. The calculated results of the simulated RMSA parameters are presented and explained in details in the following sub sections.

3.1. RETURN LOSS (S11) AND VSWR

The computed results of the S11 parameters for the single element and (1x2) array element RMSA are shown in Fig. 4. It is clearly seen from this figure that the simulated (S11) for single element reaches a value of (-34.91) dB with a bandwidth of (85) MHz in HFSS, while it arrives a value of (-44.97) dB and bandwidth of (92) MHz at (3.6) GHz with implementation of CST. While, the values of (S11) are (-17.15 and -17.58) dB for (1x2) elements RMSA arrays within a bandwidth of (200 and 204.42) MHz for HFSS and CST, respectively. However, the value of S11 parameter as a function of frequency for (2x2) and (1x4) array types are (-17.55 and -27.58) dB with a bandwidth (229.49 and 215) MHz and (-14.014 and -12.969) dB with a bandwidth (129.12 and 106.90) MHz by implementing CST and HFSS, respectively as demonstrated in Fig. 5. These computed results reveal a significant enhancement in the bandwidth with increasing antenna array elements and values of (S11) are more less than (-10) dB, which indicate the impedance matching as well [9,14]. Additionally, these results also displays that the (2x2) RMSA array configuration provide higher bandwidth operation compare to the other array arrangement considered in this investigation.

In contrast, the (VSWR) that is represent the amount of radio frequency power that properly converted from the source into the transmission line towards a load is also computed for all considered cases and with both simulation methods [31]. The calculated results of VSWR obtained for single element and (1x2) array RMSA are (1.045 and 1.011) dB and (1.305 and 1.325) with the implementation of CST and HFSS, respectively as shown in Fig. 6. Besides, the simulated results of VSWR predicted by both CST and HFSS techniques for (2x2) and (1x4) array types are, respectively, (1.094 and 1.350) dB and (1.579 and 1.945) dB as demonstrated in Fig. 7. Accordingly, these figures indicates that the VSWR for all considered cases are smaller than 2 ($VSWR < 2$) which

is regarded as an optimal limit which leads to enhanced antenna radiation performance for use in S-band wireless communication systems [31].

3.2. ANTENNA RADIATION PATTERN

The antenna radiation pattern define as the graphical descriptions of radiation fields and through which the characteristics of antenna radiation power would be identified over large distances and in different spatial orientations [32]. Figs. (8 and 9), display the three-dimensional calculated gain, consequently for the proposed single element, (1x2), (2x2) and (1x4) RMSA arrays. The computed results achieve by HFSS simulator as depicted in Figs 8(a) and 9(a) suggest that the maximum values of gain are (8.68, 10.35, 10.43 and 10.52) dB for advanced single element, (1x2), (2x2) and (1x4) (RMSA arrays operating at (3.6) GHz, respectively.

Furthermore, the higher antenna gains values provided by CST techniques for single element, (1x2), (2x2) and (1x4) RMSA arrays are, respectively, (7.36, 9.8, 9.87 and 10.30) dB, as obviously seen in Figs. 8(b) and 9(b). Generally, according to the results displayed in these figures, one clearly observes that the antenna gains increase with increasing antenna elements and all of which lies within an acceptable value reliable for S-band application systems. In addition, these results also implies that the (1x4) and (2x2) array configura-

tions maintain higher antenna gain compare to the other considered geometrical arrangements. Besides, both simulation techniques are also employed to compute the two-dimensional polar pattern view of E-plane radiation fields and the results are graphically presented in Figs. 10 for single element, (1x2), (2x2) and (1x4), respectively. It is clearly seen from these figures that as the number of patch element increase more side and back lobes are introduced while maintain the main beam in the same direction. Among which, the (2x2) array configuration provide lower side lobe level power as can be depicted in Fig 10(c) which support the reliability of this configuration to be applicable for more practical wireless communication systems.

Moreover, the overall RMSA array parameters such as, S11, VSWR, gain, directivity, efficiency and bandwidth that have been achieved with both simulation softwares and for each considered array configurations operating at (3.6 GHz) are summarized in Table 2.

Finally, these calculated parameters are compared to their corresponding value achieved by other researcher working on the same array configuration and operating at various S-band frequencies and the results are presented in Table 3. The accuracy and advancement of our calculated gain and bandwidth parameters are observed from this table as compared to the corresponding available works performed previously.

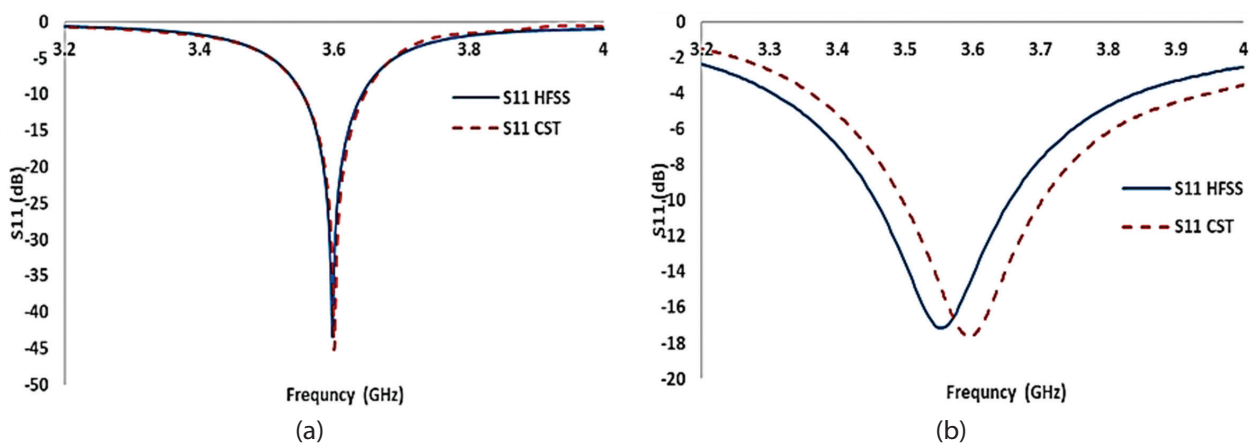


Fig. 4. Change of S11 depended on frequency for (a) single element and (b) (1x2) array

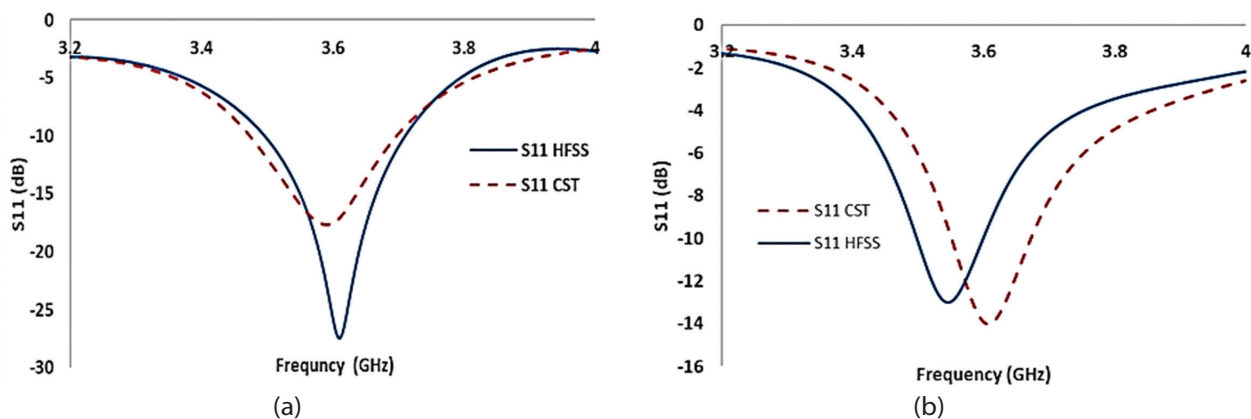


Fig. 5. Change of S11 depended on frequency for (a) (2x2) and (b) (1x4) array

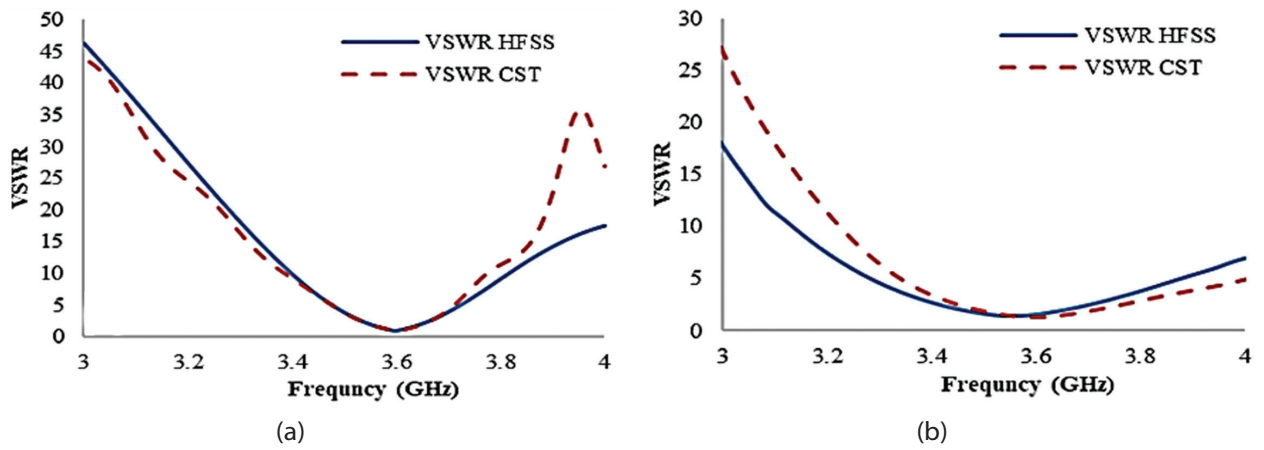


Fig. 6. Change of VSWR depended on frequency for (a) single element and (b) (1x2) array

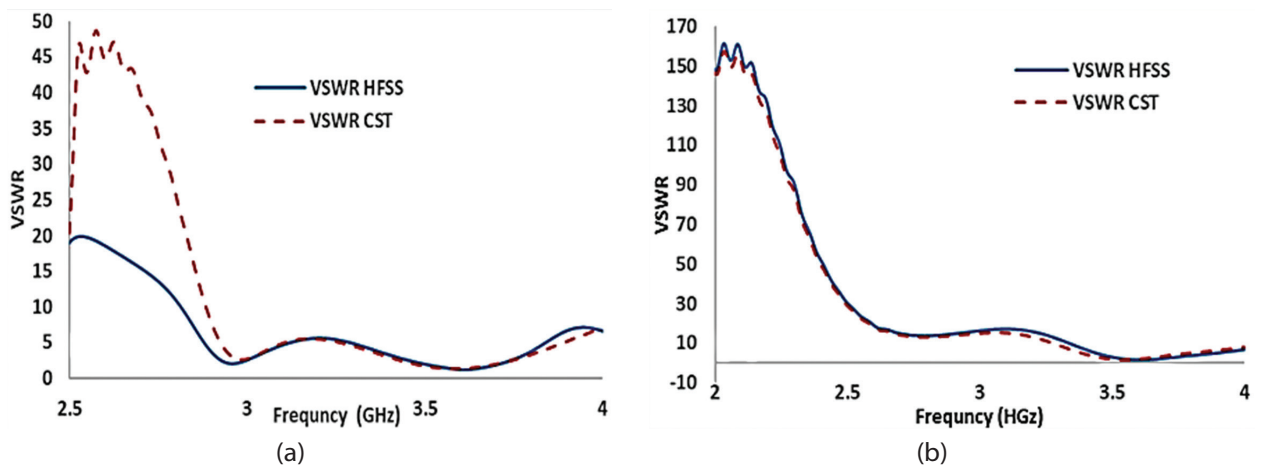


Fig. 7. Change of VSWR depended on frequency for (a) (2x2) and (b) (1x4) array

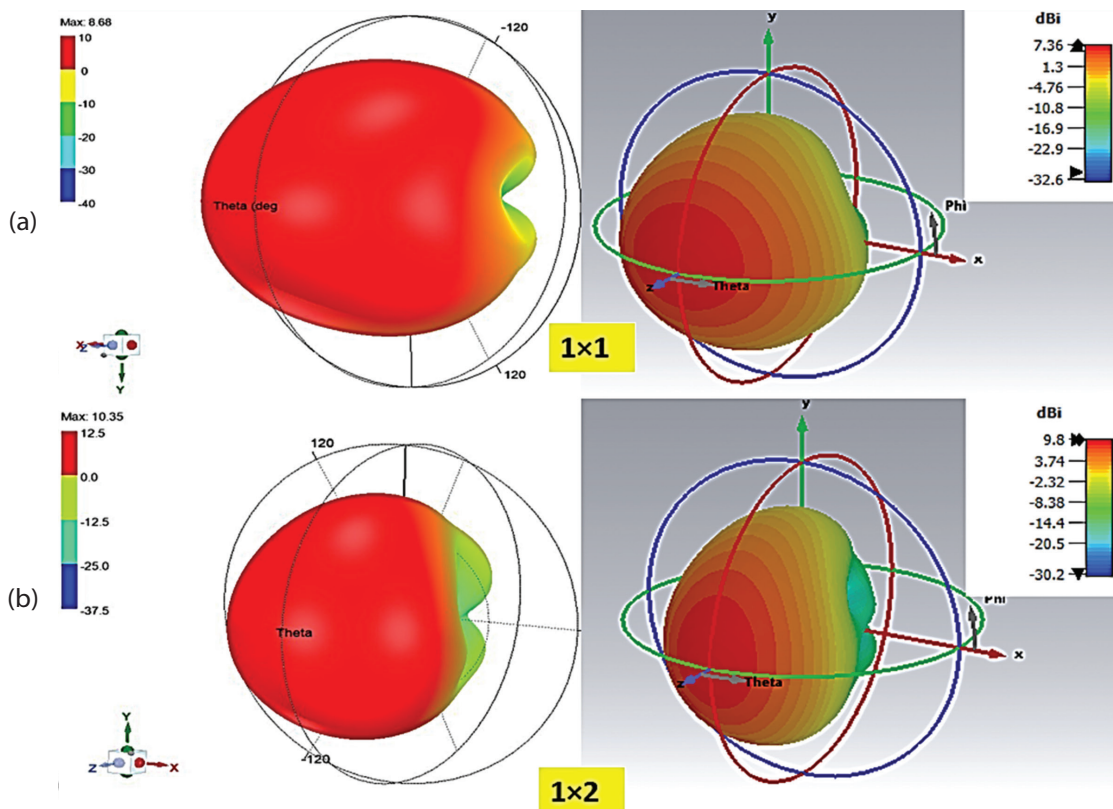


Fig. 8. 3D view of RMSA gain by (a) HFSS and (b) CST for single element and (1x2) arrays

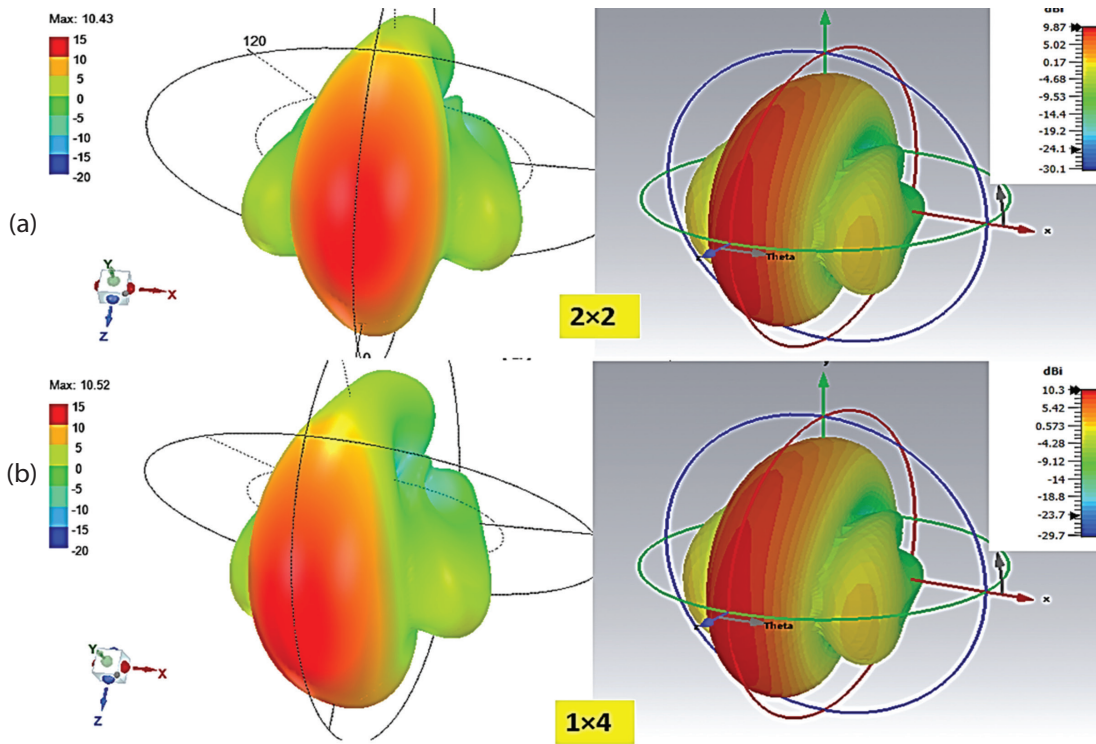


Fig. 9. 3D view of RMSA gain by (a) HFSS and (b) CST for (2x2) and (1x4) arrays

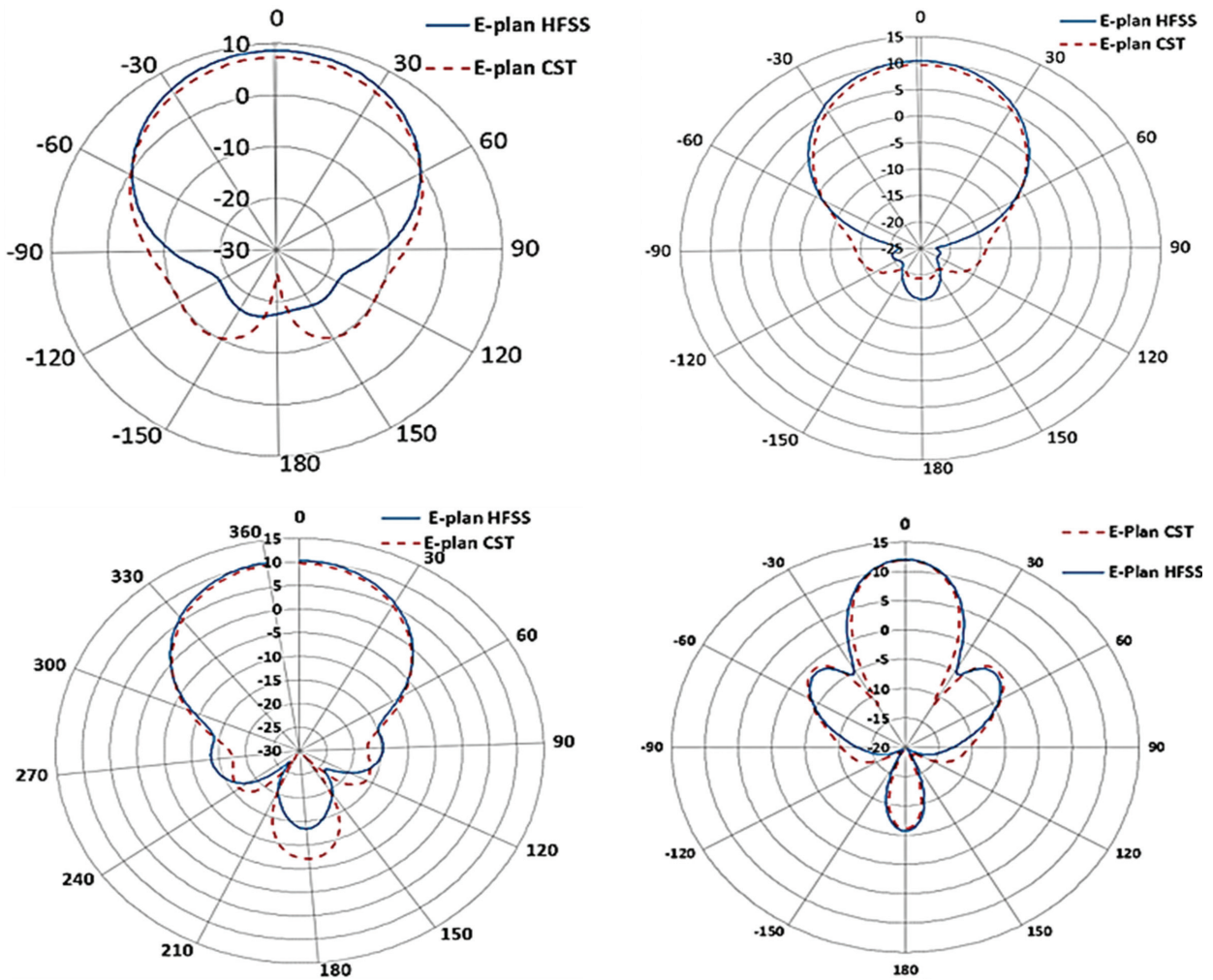


Fig. 9. 2D polar view of E-plane for (a) single element (b) (1x2) (c) (2x2) and (d) (1x4) arrays

Table 2. Overall simulated antenna chrematistic results for each considered arrays.

Element No.	Simul. Tech.	S11(dB)	VSWR	Gain (dB)	Dir(dB)	Efficiency %	BW(MHz)
Single	HFSS	-33.12	1.045	8.61	8.73	98.62	85.00
1×2	HFSS	-17.15	1.325	10.35	10.44	99.13	200.00
2×2	HFSS	-27.42	1.094	10.43	10.81	99.42	215.00
1×4	HFSS	-12.96	1.945	10.52	10.49	99.71	106.90
Single	CST	-44.97	1.011	7.36	8.30	88.67	92.28
1×2	CST	-17.58	1.304	9.80	10.10	97.03	204.42
2×2	CST	-17.55	1.305	9.87	10.30	97.19	229.49
1×4	CST	-14.01	1.579	10.30	10.70	95.82	129.12

Table 3. Comparison of propose antenna parameter results with previous researches.

No. Elements	S11(dB)	VSWR	Gain (dB)	Dir(dB)	BW(MHz)	f (MHz)	Ref.
Single Element (FEM)	-15.62	1.440	6.09	7.09	-	2.40	[20]
	-16.624	-	5.1877	-	55	2.40	[29]
	-22.89		3.29		70	2.38	[30]
	-22.417	1.317		6.9981	20	1.44	[31]
	-18.5		1.26	-	-	2.40	[13]
	-33.12	1.045	8.610	8.660	85.00		Present work
(1×2) Elements (FEM)	-17.30	1.670	7.58	8.14	-	2.40	[20]
	-9.5	-	9.186	-	-	2.40	[29]
	-11.5	-	4.17	-	15	2.50	[30]
	-22.99	1.150	-	4.62	250	5.50	[32]
	-27.61	1.080	-	7.35	205	5.55	[32]
	-7.25	2.540	9.24	-	-	2.40	[13]
	-19.31	1.650	11.42	12.17	-	2.40	[20]
-17.15	1.325	10.310	10.408	200.00		Present work	
(2×2) Elements (FEM)	-30.34	1.070	-	12.91	173	5.58	[32]
	-32.34	1.170	-	11.28	186	5.55	[32]
	-72.5555	0.004	-	14.739	20	1.44	[31]
	-27.42	1.094	10.430	10.450	215.00		Present work
(1×4) Elements (FEM)	-22	-	13.2 -	--	60	2.40	[29]
	-19.906	1.762		14.576	20	1.44	[31]
	-8.25	2.260	10.29	-	-	2.40	[13]
	-12.969	1.945	10.52	12.250	106.90		Present work
Single Element (FIT)	-13.934	1.502	2.144	-	61	2.20	[26]
	-44.830	-	1.369	6.759	-	2.40	[21]
	-12.89	1.590	4.25	5.63	-	2.40	[33]
	-26.5	-	4.2	-	190	3.50	[34]
	-28.76		2.52	2.60	130	3.60	[3]
	-44.97	1.011	7.357	8.298	92.28		Present work
(1×2) Elements (FIT)	-20.371	1.211	4.58	-	83	2.2	[26]
	-11.21	1.750	9.225	9.633	29.3	2.45	[25]
	-17.58	1.304	9.670	9.924	204.42		Present work
(2×2) Elements (FIT)	-18.086	1.280	4.714	-	83.6	2.2	[26]
	-17.55	1.305	9.87	11.90	229.49		Present work
(1×4) Elements (FIT)	-20.372		5.284	10.59	-	2.5	[21]
	-7.95	2.530	7.31	9.58	-	2.4	[33]
	-12.74	1.590	11.73	11.79	35.5	2.45	[25]
	-22.5	-	4.52	-	-	3.5	[12]
	-14.014	1.579	10.30	12.30	129.12		Present work

4. CONCLUSIONS

This article has proposed a theoretical calculation methodology for designing a new single element, 1x2, 2x2 and 1x4 elements RMSA array operating at (3.6 GHz) suitable for S-band wireless communication systems. The array performances presented here were investigated using both HFSS and CST numerical methods. Generally, the computed results reveals that a lower side lobe level power, reasonable gain and wider bandwidth of the order of (10.43, 10.40) dB and (215, 229.49) MHz was achieved, respectively, for the case of (2x2) array elements. In addition, the simulation results indicated that the values of S11 are (-34.91, -17.15, -27.42 and -12.26) dB with bandwidth (85.00, 200.00, 215 and 106.4) MHz and gain of (8.68, 10.35, 10.43 and 7.86) dB by the execution of HFSS for designed single element, (1x2), (2x2) and (1x4) array configurations, respectively. Whereas the corresponding values of S11 are (-44.97, -17.58, -17.55 and -14.01) dB, bandwidth (92.28, 204, 229.49 and 129.12) MHz and gain of (7.36, 9.8, 10.40 and 9.87) dB with the implementation of CST for suggested element arrays, respectively. The issue of low bandwidth and gain which usually achieved by single element RMSA array has been resolved by the novel (RMSA) array models provided in this paper. The newly developed antennas are suitable to employ in radar, WLAN, S-band communications, and 5G wireless communication technologies.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Financing

The authors received no specific funding for this work.

5. REFERENCES

- [1] Z. Ying, "Antennas in cellular phones for mobile communications", *Proceedings of the IEEE*, Vol. 100, No. 7, 2012, pp. 2286-2296.
- [2] S. K. Ezzulddin, S. O. Hasan, M. M. Ameen, "Microstrip patch antenna design, simulation and fabrication for 5G applications", *Simulation Modelling Practice and Theory*, Vol. 116, 2022, p. 102497.
- [3] A. Kapoor, R. Mishra, P. Kumar, "Compact wideband-printed antenna for sub-6 GHz fifth-generation applications", *International Journal on Smart Sensing and Intelligent Systems*, Vol. 13, No. 1, 2020, pp. 1-10.
- [4] K. Gilly, C. Bernad, P. J. Roig, S. Alcaraz, S. Filiposka, "End-to-end simulation environment for mobile edge computing", *Simulation Modelling Practice and Theory*, Vol. 121, 2022, p. 102657.
- [5] R. Cárdenas, P. Arroba, R. Blanco, P. Malagón, J. L. Risco-Martín, J. M. Moya, "Mercury: A modeling, simulation, and optimization framework for data stream-oriented IoT applications", *Simulation Modelling Practice and Theory*, Vol. 101, 2020, p. 102037.
- [6] S. N. Dahake, N. R. Kolhare, "Design & development of 2x2 microstrip patch antenna array in SAR for satellite application", *Proceedings of the International Conference on Computing Methodologies and Communication*, 2017, pp. 800-804.
- [7] N. Ab Wahab, S. A. Nordin, W. N. W. Muhamad, S. S. Sarnin, "Microstrip Rectangular Inset-Fed Patch Array Antenna for WiMax Application", *Proceedings of the IEEE International RF and Microwave Conference*, 2020, pp. 1-4.
- [8] M. M. Nahas, "Gain Enhancement of a Microstrip Patch Antenna with Quasi H-Shaped Slot for UHF RFID Reader", *Journal of Applied Science and Engineering*, Vol. 26, No. 4, 2022, pp. 565-569.
- [9] M. Nahas, "Design of a high-gain dual-band LI-slotted microstrip patch antenna for 5G mobile communication systems", *Journal of Radiation Research and Applied Sciences*, Vol. 15, No. 4, 2022, p. 100483.
- [10] H. Huang, W. Zhao, "A triple-band multi-functional antenna for LTE/GSM/UMTS/WiMAX handsets", *Proceedings of the Asia-Pacific Microwave Conference*, Nanjing, China, December 2015, Vol. 2, pp. 1-3.
- [11] M. Z. B. Chowdhury, M. T. Islam, H. Rmili, I. Hossain, M. Z. Mahmud, M. Samsuzzaman, "A low-profile rectangular slot antenna for sub-6 GHz 5G wireless applications", *International Journal of Communication Systems*, Vol. 35, No. 17, 2022, p. e5321.
- [12] Y. S. Khraisat, "Design of 4 elements rectangular microstrip patch antenna with high gain for 2.4 GHz applications", *Modern Applied Science*, Vol. 6, No. 1, 2012, p. 68.
- [13] A. B. Obot, G. A. Igwue, K. M. Udofia, "Design and simulation of rectangular microstrip antenna arrays for improved gain performance", *International Journal of Networks and Communications*, Vol. 9, No. 2, 2019, pp. 73-81.
- [14] A. Santoso, F. Cahyono, I. Suwondo, B. Harianto, "Microstrip Antenna Design with Array Rectangular Patch

- 2x2 for Ship Radar at 2.2 GHz", *Journal of Physics: Conference Series*, Vol. 2117, No. 1, 2021, p. 012019.
- [15] B. S. S. Reddy, P. Parvathi, B. S. Kumar, "Construction and Implementation of Coaxial Probe Feed Microstrip Patch Array Antenna", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 8, No. 11, 2019, pp. 814-821.
- [16] V. R. Ekke, P. L. Zade, "Design and Implementation of T-junction Triangular Microstrip Patch Antenna Array for Wireless Applications", *International Journal of Engineering and Technology*, Vol. 8, No. 5, 2016, pp. 2105-2114.
- [17] M. M. Alam, R. Azim, N. M. Sobahi, A. I. Khan, M. T. Islam, "A dual-band CPW-fed miniature planar antenna for S-, C-, WiMAX, WLAN, UWB, and X-band applications", *Scientific Reports*, Vol. 12, No. 1, 2022, p. 7584.
- [18] F. Wang et al. "Design of an ultra-thin absorption layer with magnetic materials based on genetic algorithm at the S band", *Journal of Magnetism and Magnetic Materials*, Vol. 451, 2018, pp. 770-773.
- [19] S. O. Hasan, "Design and Comparison Study of Circular and Elliptical Microstrip Patch Antennas for 5G Applications", *International Journal on Communications Antenna and Propagation*, Vol. 12, No. 6, 2022, pp. 437-447.
- [20] J.-M. Jin, D. J. Riley, "Finite element analysis of antennas and arrays", John Wiley & Sons, 2008.
- [21] T. Weiland, "A discretization model for the solution of Maxwell's equations for six-component fields", *Archiv Elektronik und Uebertragungstechnik*, Vol. 31, 1977, pp. 116-120.
- [22] M. J. Hakeem, M. M. Nahas, "Improving the Performance of a Microstrip Antenna by Adding a Slot into Different Patch Designs", *Engineering, Technology & Applied Science Research*, Vol. 11, No. 4, 2021, pp. 7469-7476.
- [23] A. Abdelaziz, E. K. I. Hamad, "Design of a Compact High Gain Microstrip Patch Antenna for Tri-Band 5 G Wireless Communication", *Frequenz*, Vol. 73, No. 1-2, 2019, pp. 45-52.
- [24] Y. Huang, "Antennas: From Theory to Practice", 2nd Edition, John Wiley & Sons, Ltd, 2021.
- [25] L. C. Yu, M. R. Kamarudin, "Investigation of patch phase array antenna orientation at 28 GHz for 5G applications", *Procedia Computer Science*, Vol. 86, 2016, pp. 47-50.
- [26] R. J. Mailloux, "Phased array antenna handbook", Artech house, 2017.
- [27] C. A. Balanis, "Antenna Theory Analysis and Design", 4th Edition, John Wiley & Sons, 2016.
- [28] C. A. Balanis, "Antenna Theory Analysis and Design", 3rd Edition, Wiley-Interscience, 2005.
- [29] T. T. S. Borel, A. R. Yadav, U. Shah, "Design of rectangular patch array antenna for satellite communication", *Proceedings of the 3rd International Conference on Computing Methodologies and Communication*, Erode, India, 27-29 March 2019, pp. 759-764.
- [30] V. Prakasam, P. Sandeep, K. A. LaxmiKanth, "Rectangular micro strip patch array antenna with corporate feed network for wireless communication applications", *Proceedings of the 5th International Conference on Communication and Electronics Systems*, Coimbatore, India, 10-12 June 2020, pp. 311-316.
- [31] O. Sokunbi, H. Attia, S. I. Sheikh, "Microstrip antenna array with reduced mutual coupling using slotted-ring ebg structure for 5G applications", *Proceedings of the IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, Atlanta, GA, USA, 7-12 July 2019, pp. 1185-1186.
- [32] R. Kazemi, S. Yang, S. H. Suleiman, A. E. Fathy, "Design procedure for compact dual-circularly polarized slotted substrate integrated waveguide antenna arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 67, No. 6, 2019, pp. 3839-3852.
- [33] M. T. Srilakshmi, S. T. Bhanu, "Performance Comparison of S Band Antenna with Series Fed and Corporate Fed Microstrip Array", *International Journal of Engineering & Technology*, Vol. 7, No. 2, 2018, p. 4.
- [34] W. N. Ab Wahab, W. Muhamad, Z. I. Khan, S. S. Sarnin, "Microstrip array antenna with inset-fed for WLAN application", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 17, No. 1, 2020, pp. 340-346.

Noise Effects on a Proposed Algorithm for Signal Reconstruction and Bandwidth Optimization

Original Scientific Paper

Ahmed F. Ashour

Electrical and Computer Engineering Department,
Idaho State University
921 S 8th Ave 83201, Pocatello, USA
ahmedashour@isu.edu

Ashraf A. M. Khalaf

Department of Electronics and Communications
Engineering,
Faculty of Engineering, Department of Computer
Science, Minia University
Damaris 61519, El-Minia, Egypt
ashraf.khalaf@mu.edu.eg

Aziza I. Hussein

Electrical and Computer Engineering Department,
Effat University
Al-Nazlah Al-Yamaniyah 34689, Jeddah, KSA
azibrahim@effatuniversity.edu.sa

Hesham F. A. Hamed

Faculty of Engineering, Egyptian Russian University
Faculty of Engineering, Minia University
Badr 11829, Cairo, Egypt
Damaris 61519, El-Minia, Egypt
hesham.fathy@eru.edu.eg

Ashraf Ramadan

Electrical and Computer Engineering Department,
Higher Technological Institute
Industrial Area2 44629, 10th of Ramadan, Egypt
ashraf.ramadan@hti.edu.eg

Abstract – The development of wireless technology in recent years has increased the demand for channel resources within a limited spectrum. The system's performance can be improved through bandwidth optimization, as the spectrum is a scarce resource. To reconstruct the signal, given incomplete knowledge about the original signal, signal reconstruction algorithms are needed. In this paper, we propose a new scheme for reducing the effect of adding additive white Gaussian noise (AWGN) using a noise reject filter (NRF) on a previously discussed algorithm for baseband signal transmission and reconstruction that can reconstruct most of the signal's energy without any need to send most of the signal's concentrated power like the conventional methods, thus achieving bandwidth optimization. The proposed scheme for noise reduction was tested for a pulse signal and stream of pulses with different rates (2, 4, 6, and 8 Mbps) and showed good reconstruction performance in terms of the normalized mean squared error (NMSE) and achieved an average enhancement of around 48%. The proposed schemes for signal reconstruction and noise reduction can be applied to different applications, such as ultra-wideband (UWB) communications, radio frequency identification (RFID) systems, mobile communication networks, and radar systems.

Keywords: signal reconstruction, bandwidth optimization, AWGN, noise reduction, baseband signal, NMSE, noise reject filter

1. INTRODUCTION

Every day, current analog signals like our voices are used in the real world. These signals need to be sampled, quantized, and encoded before being processed in a digital format [1]. "Reconstruction" or "interpolation" of the signal refers to the process of returning the sampled signal to its original analog form. Every one of our everyday devices undergoes the reconstruction process, which is crucial for restoring the signal to its

original form [2]. Multiple schemes have been proposed for reconstructing a bandlimited signal from irregularly spaced sampling data [3-5]. The phase of the signal can be used for reconstructing the signal [6], and other schemes can be used without using it [7].

There are multiple research studies that have developed various strategies for reconstructing pulse and stream of pulse signals and used them in various systems. A proposed effective doppler-based signal reconstruc-

tion technique is used to restore the doppler spectrum in synthetic aperture radar (SAR) [8]. Other schemes based on compressed sensing (CS) are used for reconstructing the electrocardiogram (ECG) pulse signals [9-11], while [12] suggested a deep learning (DL)-based CS technique to reconstruct the ultrawideband (UWB) signal. Another approach is presented to the fast and accurate nonlinear pulse signal reconstruction for electromagnetic (EM) sensors and its applications [13]. Other algorithms are used for reconstructing variable-width pulses [14] and ultra-short pulses using DL [15]. The reconstruction of a stream of pulses for ultrasonic imaging was proposed using a noiseless finite rate innovation (FRI) method [16,17]. Many low-pass reconstruction problems have been successfully resolved via extrapolation [18-20].

In such real-world scenarios, it is necessary to establish appropriately accurate estimates of the inverse Fourier transform (IFT) based on this imperfect knowledge. The primary objective of the prior proposed work in [21] is to present a new method for extrapolating a finite-frequency segment of a time-limited baseband pulse signal. The approach relies on choosing a starting and ending frequency (f_s and f_e) to send through the bandlimited channel while utilizing a selective band-pass filter (BPF). A suggested algorithm is provided and put into practice on the receiver side in order to reconstruct the signal once more and recover the majority of the signal's energy. By adjusting the adaptive band-reject filter (BRF) in the receiver to only select the same transmitted band limits from f_s to f_e to reconstruct the signal, we can optimize the use of the channel bandwidth instead of selecting the entire signal spectrum or the main lobe band that contains the majority of the signal's power, as required by conventional methods. The results of the proposed algorithm showed that the algorithm converges when the number of iterations of the reconstruction algorithm is increased, as in the case of working with a noiseless channel.

This paper extends the work in [22] by studying the effect of additive white Gaussian noise (AWGN) on our previously proposed baseband signal transmission and reconstruction scheme and proposing a new technique to minimize the effect of noise using an adaptive noise reject filter (NRF), which consists of a bandpass filter (BPF) with the same starting and ending frequency parameters as the whole system. The proposed technique has been tested on a pulse and a stream of pulses and showed good reconstruction performance in terms of the normalized mean square error (NMSE) when compared to a noise-free channel. A list of used abbreviations is shown in Table 1.

The paper is organized as follows: Section 2 introduces the signal transmission and reconstruction scheme proposed previously and applied to a pulse and stream of pulses. Section 3 discusses the traditional and proposed solutions for reducing the noise added to the channel, their limitations, and the proposed applications of the scheme. Finally, Section 4 presents the conclusion.

Table 1. List of abbreviations

Abbreviation	Meaning
6G	sixth generation
AWGN	additive white Gaussian noise
BPF	bandpass filter
BPF	band-pass filter
BRF	band-reject filter
CS	compressed sensing
DFT	discrete fourier transform
DL	deep learning
dMRI	diffusion magnetic resonance imaging
DSIC	digital self-interference cancellation
DWT	discrete wavelet transform
ECG	electrocardiogram
EM	electromagnetic
EVVM	error vector magnitude
FFT	fast Fourier transform
FRI	finite rate innovation
FT	Fourier transform
IFT	inverse Fourier transform
IoT	internet of things
KPCA	kernel principal component analysis
LPF	low-pass filter
LPR	lost pulse ratio
LS	least squares
MSE	mean squared error
NMSE	normalized mean square error
NRF	noise reject filter
OFDM	orthogonal frequency division multiplexing
PSNR	peak signal-to-noise ratio
RFID	radio frequency identification
SAR	synthetic aperture radar
SI	self-interference
SPR	spurious pulse ratio
SVD	singular value decomposition
UHF	ultra-high frequency
UWB	ultrawideband
UWOC	underwater optical wireless communication
V2V	vehicle-to-vehicle communications
WSN	wireless sensor network

2. BASEBAND SIGNAL TRANSMISSION AND RECONSTRUCTION SCHEME

2.1 INTRODUCTION

The two main signal transmission mechanisms in any communication system are baseband and bandpass transmission. The baseband signal is distinguished by its low-frequency components, which include the DC component, such as the signal of the information source, such as human speech. Modulating the baseband signal to higher frequencies results in a bandpass signal. The resultant bandpass signal is concentrated around the carrier frequency, $\pm f_c$. Figs. 1(a) and (b) show examples of baseband signals such as a pulse signal $x_p(t)$ with a duration of 2 ms and a stream of pulses $x_{sp}(t)$ with a data rate of 8 Mbps.

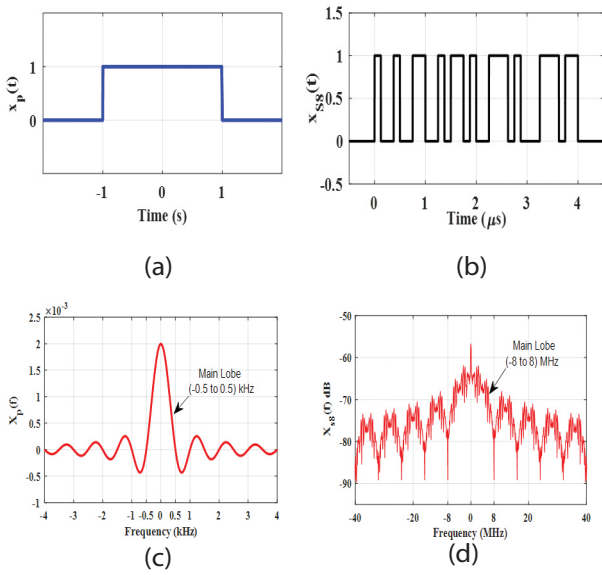


Fig. 1. Baseband signals (a) A pulse signal $x_p(t)$ with $T=2$ ms (b) A pulse stream $x_{SB}(t)$ with $R_b=8$ Mbps (c) The spectrum of $x_p(t)$ (d) The spectrum of $x_{SB}(t)$

The Fourier transform (FT) can be used to obtain the amplitude spectra of $x_p(t)$ and $x_{SB}(t)$, as shown in Figs. 1(c) and (d), respectively. The figures show that the maximum power of these signals is concentrated around the zero frequencies in the main lobe (more than 90%), and the remaining power of the signals is distributed along the side lobes. Theoretically, these signals have infinite bandwidth and cannot be transmitted over band-limited channels. So, in order to send these signals over band limited channels, a low-pass filter (LPF) is applied to limit the signals' bandwidth and then transmitted over those channels as in conventional systems.

2.2. THE TRANSMISSION TECHNIQUE

Fig. 2 depicts the block diagram of the proposed technique for transmitting and reconstructing baseband signals. The transmitter consists of a generator to generate the baseband information $x(t)$, which is a pulse or stream of pulses in our study. The transmission technique depends on selecting any window (W) of the baseband signal using an adaptive BPF instead of sending the whole signal's spectrum or the main lobe only. The boundaries of the BPF are defined by a starting frequency f_s and an ending frequency f_e . The filter will produce a signal $g(t)$, which is distorted due to losing some of the signal's spectral information.

At the transmitter, assume that the selective BPF is adjusted to send a window (0.3-3) kHz from the original pulse's spectrum $x_p(f)$, which represents 17.2% of its total average power. Also, assume that the transmitted window of the pulse stream $x_{SB}(f)$ is (0.3-8.3) MHz, which represents 47.1% of its total average power.

Then, the transmitted signals $g_p(t)$ and $g_{SB}(t)$ will seem to be distorted due to losing some energy from their spectra, as shown in Fig. 3.

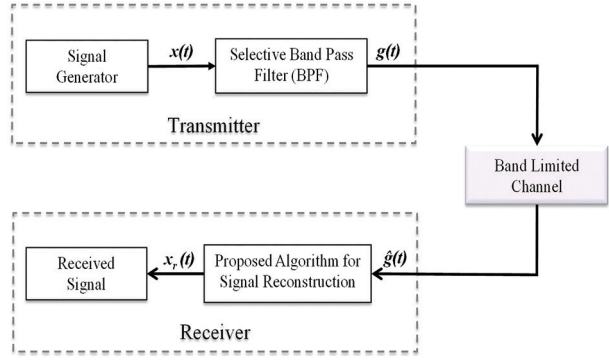


Fig. 2. Proposed technique's block diagram

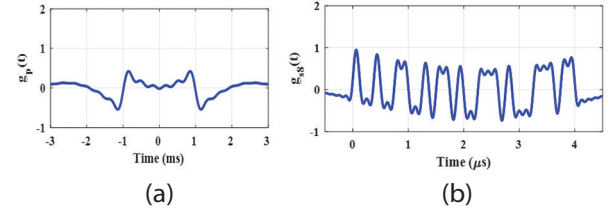


Fig. 3. Transmitted distorted signals of (a) $x_p(t)$ with $W=(0.3-3)$ kHz (b) $x_{SB}(t)$ with $W=(0.3-8.3)$ MHz

2.3. THE RECONSTRUCTION ALGORITHM

The block diagram of the baseband signal reconstruction algorithm at the receiver is shown in Fig. 4. The aim of this algorithm is to reconstruct the signal $x_r(t)$ with the knowledge of a segment $G(f)$ of the main signal's spectrum $X(f)$, then extrapolate the original signal's spectrum by making use of the received segment $\hat{G}(f)$, and a prior knowledge about the time extent of the transmitted signal.

The steps for reconstructing the baseband signal, as explained in Fig. 4, are as follows:

Step (1): IFT of the received signal's spectrum. The output is a non-time-limited signal.

Step (2): Multiplying by a gate (rect) function $p(t)$ with the same time extent as the original signal results in $s(t)$.

Step (3): Applying the FT to $s(t)$ to obtain $S(f)$, a non-bandlimited signal.

Step (4): Applying $S(f)$ to a BRF with the same starting and ending frequencies of the transmitter (f_s and f_e) to obtain $C(f)$.

Step (5): Adding the spectrum $C(f)$ with the known received signal's spectrum to be inserted into its dead space to be raised gives the first estimate of the reconstructed signal's spectrum $X_1(f)$.

Step (6): Calculating the IFT of $X_1(f)$ and repeating the loop again until reaching the required shape of the spectrum after N iterations.

Step (7): Calculating the IFT of $X_N(f)$ to get $x_n(t)$, which is the reconstructed signal after a certain number of iterations. The reconstructed baseband pulse stream signal of $x_{SB}(t)$ and its spectrum after 1 and 300 iterations are shown in Fig. 5.

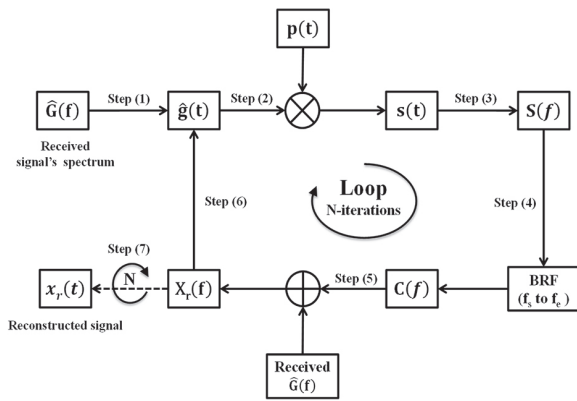


Fig. 4. The block diagram of baseband signal reconstruction algorithm

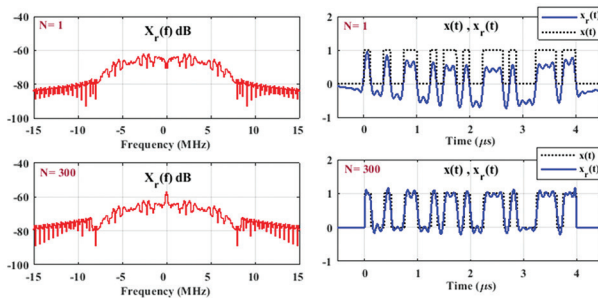


Fig. 5. The reconstructed pulse stream signal (8 Mbps) and its spectrum after 1 and 3000 iterations

3. RESULTS AND DISCUSSIONS

3.1. THE EFFECT OF AWGN ON THE ALGORITHM

In this section, the effect of adding the AWGN to the received signal will be determined, and the algorithm's behavior on the reconstructed signal will be checked. It is assumed that the transmitted pulse $g(t)$ with a transmitted window band (0.3-3) kHz is passed through an AWGN with a SNR of 10 dB. The signal with noise $\hat{g}_n(t)$ in Fig. 6 (a) is received and applied to the proposed algorithm in Fig. 4 to reconstruct the baseband pulse signal again. As seen in Figs. 6 (b), (c), and (d), increasing N results in increasing the noise in the reconstructed signal. This is due to the accumulated addition of the received signal spectrum with noise $\hat{G}_n(f)$ in step 5 of the proposed algorithm shown in Fig. 4. So, the noise effect is increased by increasing the number of iterations; thus, the algorithm diverges.

Table 2 presents the impact of changing three key parameters – transmitted window bandwidth (W), number of iterations (N), and signal-to-noise ratio (SNR) – on the reconstruction status in the presence of AWGN.

Each column represents the variation of a single parameter, while the other parameters are held constant. The first column demonstrates the effect of altering the transmitted window bandwidth while keeping N and SNR fixed. Increasing W leads to a decrease in NMSE, indicating an improvement in the reconstruction's performance. The second column illustrates the effect of

adjusting N while keeping W and NMSE constant. Here, we observe that increasing the number of iterations can result in a higher NMSE value, which suggests a negative impact on the reconstruction accuracy. The last column shows the effect of modifying SNR while keeping NMSE and N constant. It is evident that raising the SNR can lower the NMSE value, thereby improving the reconstruction performance. Therefore, an optimal balance between these parameters can be struck to meet the desired NMSE as per the system specifications.

Table 2. Effect of Transmitted Window Bandwidth, Number of Iterations, and SNR on Reconstruction Performance in the Presence of AWGN

Window BW (W) ($N=5$, SNR=25 dB)		Variable (N) ($W=0-4$ MHz, SNR= 25 dB)		Variable (SNR) ($W=0-4$ MHz, $N=5$)	
P %	NMSE	N	NMSE	SNR	NMSE
$W=(0-2)$ MHz		5	0.0868	15 dB	0.3518
86.94	0.1389	10	0.1701	20 dB	0.1516
93.09	0.0903	15	0.3180	25 dB	0.0863
$W=(0-6)$ MHz		20	0.4714	30 dB	0.0689
95.46	0.0665				
96.48	0.0576				

So, the traditional solutions to overcome the effect of noise in traditional systems are to increase the SNR at the receiver to overcome the noise signal's power, or to increase the transmitted window band. Although these solutions may work, we will propose another effective solution to reduce the effect of noise without any need to increase the SNR at the receiver or increase the transmitted window band.

3.2. THE PROPOSED NOISE REDUCTION SCHEME

As seen in the previous sub-section, adding the AWGN to the signal causes the algorithm to act very badly, and the signal cannot be reconstructed because the noise increases with every iteration, as shown in Fig. 6, causing the reconstructed signal to be loaded with more and more noise. The block diagram of the proposed technique to reduce the noise effect on our algorithm is shown in Fig. 7. The signals that explain the proposed scheme are shown in Fig. 8.

The steps of the proposed technique are as follows:

Step 1: Calculating the FT of the received signal with noise $\hat{g}_n(t)$ to get $\hat{G}_n(f)$ as shown in Fig. 8 (a) and (b).

Step 2: Applying $\hat{G}_n(f)$ to a NRF, which consists of a BPF with the same f_s and f_e as the transmitter and the receiver, in order to reject all the noise that exists along the signal's spectrum and keep the remaining spectrum as it is to result in a less noisy signal spectrum $\hat{G}(f)$ with zero noise in the shown regions (A, B, and C) in Fig. 8(c).

Steps (3-8): Applying the resultant spectrum of step 2 [$\hat{G}(f)$] to our proposed algorithm for N iterations, such as steps (1-6) in Fig. 4, to get $XN(f)$.

Step 9: Calculating the IFT of $XN(f)$ to get the reconstructed signal $x_r(t)$ after noise reduction, as shown in Fig. 8(d).

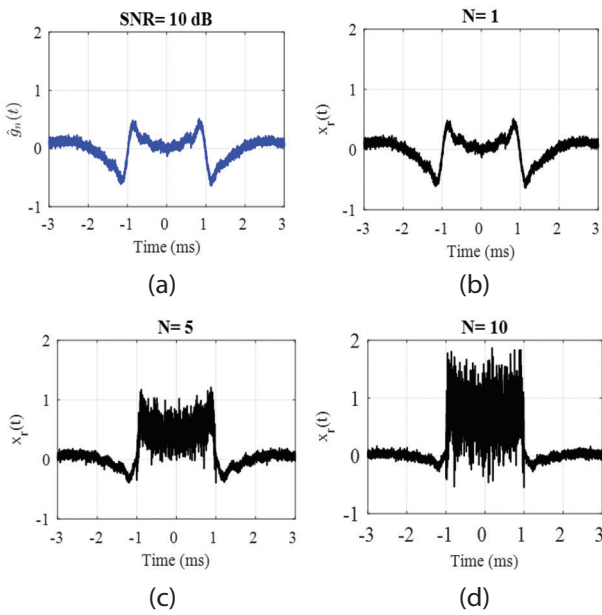


Fig. 6. Effect of increasing N in the presence of AWGN with SNR= 10 dB on the reconstructed pulse after (1, 5, and 10) iterations with a transmitted window band (0.3-3) kHz

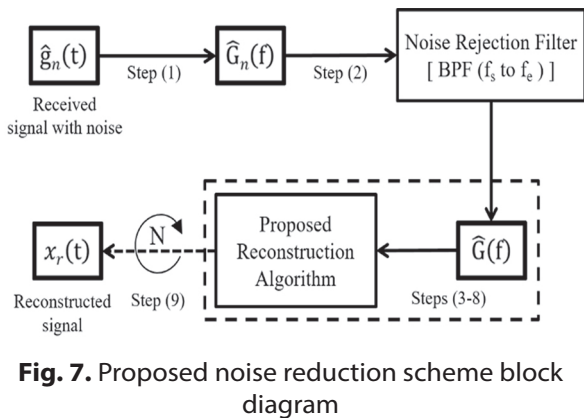


Fig. 7. Proposed noise reduction scheme block diagram

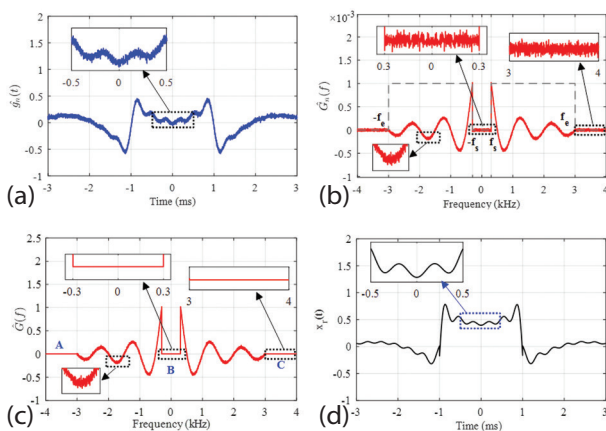


Fig. 8. Proposed noise reduction scheme signals: (a) Received signal with noise (b) FT of the received signal with noise (c) Received signal spectrum after noise reduction (d) Reconstructed signal after 5 iterations.

The proposed noise reduction technique is applied to different streams of pulses with different data rates, but we will focus on the results at 8 Mbps to be brief. Fig. 9 shows the reconstruction signals and their spectra in a noise-free channel and an AWGN for an 8 Mbps data rate with a transmitted window band (0.3-8.3) MHz, which contains $\approx 47\%$ of the total signal's average power at SNR = 10 dB.

The Fig. 8 shows the reconstructed signal with accumulated noise after 5 iterations. The middle figure shows the reconstructed stream in a noiseless channel after applying the proposed noise reduction technique. It shows a good reconstruction status of the noisy signal after applying the scheme, which is very similar to the signal that is constructed without the existence of noise. A comparison of the spectra of all these signals is shown in the figure below. Also, the spectra of the reconstructed signals with and without noise are almost identical.

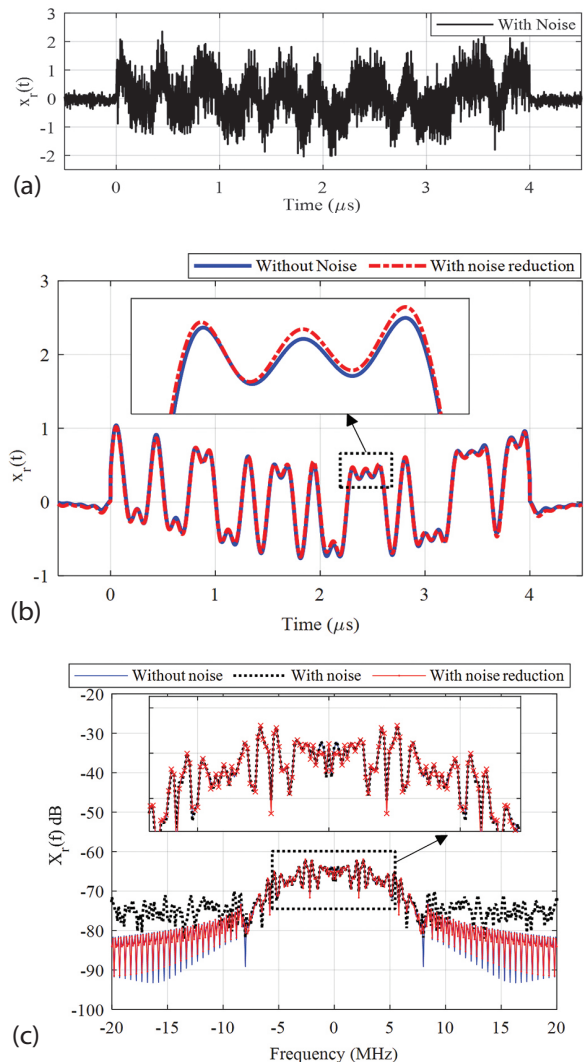


Fig. 9. Effect of noise reduction scheme on a transmitted window (0.3-8.3) MHz of an 8 Mbps stream with SNR = 10 dB, $N = 5$ (a) Reconstructed signal without noise reduction (b) Reconstructed signal with noise reduction (c) A comparison of reconstructed signals' spectra

A comparison between the NMSE of the reconstructed stream of pulses in a noise-free channel and the AWGN channel after noise reduction with the same parameters as in the previous section for different SNRs is shown in Fig. 10. For example, let us compare the reconstructed signal without noise and with 10 dB AWGN. Their NMSEs at different iterations are recorded in Table 3 for comparison. The figure also shows that when $N < 200$ for any SNR value, the reconstructed signal after reducing the noise has a better reconstruction status than the reconstructed signal when the channel is noiseless, as it has a better NMSE value of 0.0719 than the other one (0.1256). This is approved in Fig. 11(a), as the mid-level amplitude of $x_r(t)$ with noise reduction has the same mid-level amplitude of the original signal $x(t)$, but $x_r(t)$ without noise has a shifted value (0.24) by $\approx 50\%$ from $x(t)$.

At $N = 184$, there is an intersection between the two NMSE curves at 0.0723, in which the mid-level amplitudes of the reconstructed signals in both cases have an amplitude offset of 20% away from $x(t)$, as shown in Fig. 11(b). At $N = 300$, $x_r(t)$ without noise has a better performance than the other one (with noise), as it has a lower NMSE value (0.0556) and has the same mid-level amplitude value of $x(t)$ as seen in Fig. 11(c). So, it is recommended to use the proposed technique for noise reduction to reconstruct an 8 Mbps stream signal in a noisy channel at $N = 100$ or 150.

The effect of changing both N and SNR on the reconstructed signal after applying the noise reduction technique in the presence of the AWGN channel is shown in Fig. 12. It depicts that increasing N results in enhancing the NMSE value for any SNR. But fixing N while increasing the SNR value does not enhance the NMSE very well except at low SNR values (≤ 5 dB). So if the designer needs to improve the reconstruction performance using the proposed noise reduction scheme, he/she does not need to increase the signal's power, but he/she should increase the number of iterations to achieve a good reconstruction status.

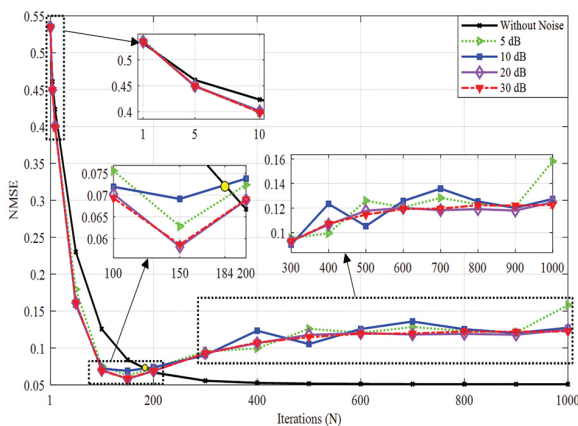


Fig. 10. A Comparison between the NMSE of the reconstructed 8 Mbps stream of pulses with a transmitted window (0.3-8.3) MHz without and with the addition of AWGN in different SNRs

Table 3. NMSE of the reconstructed 8 Mbps stream in noise-free and AWGN channels at different iterations

Reconstructed signal $x_r(t)$	NMSE at SNR= 10 dB		
	N = 100	N = 184	N = 300
Without noise	0.1256	0.0723	0.0556
With noise reduction	0.0719	0.0723	0.0902

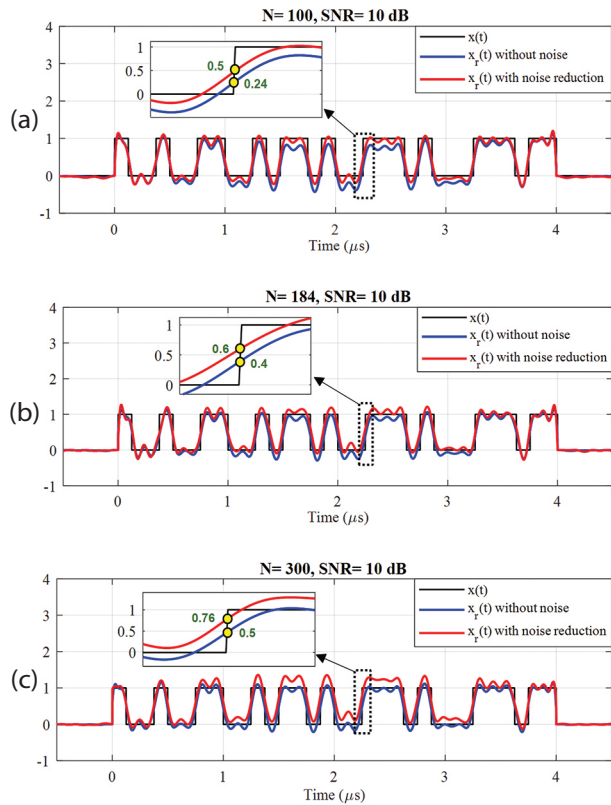


Fig. 11. A comparison between the generated 8 Mbps stream with the reconstructed streams in noise-free channel and AWGN channel at SNR= 10 dB when (a) $N = 100$, (b) $N = 184$, and (c) $N = 300$.

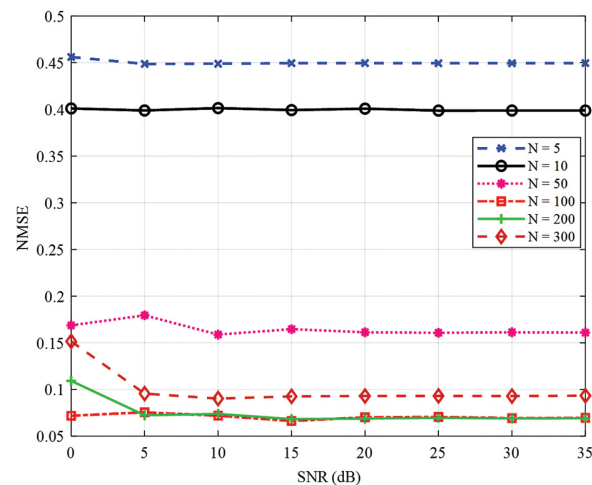


Fig. 12. Effect of changing the SNR on the reconstruction algorithm in the presence of AWGN for the same data in Fig. 9 for different iterations

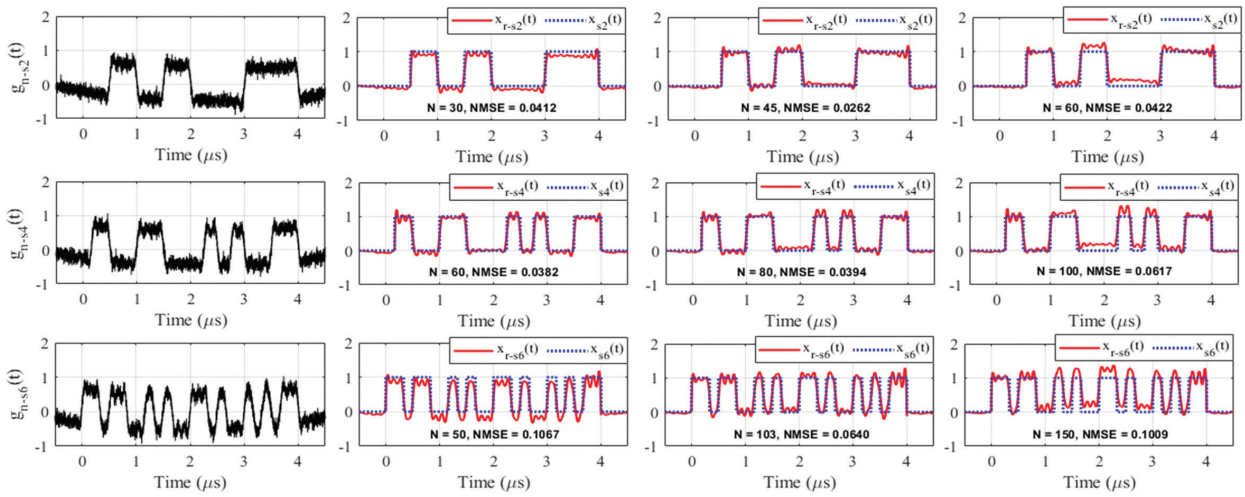


Fig. 13. Reconstructed 2, 4, and 6 Mbps data streams with noise at SNR= 10 dB, W= 0.3-8.3 MHz and different iterations

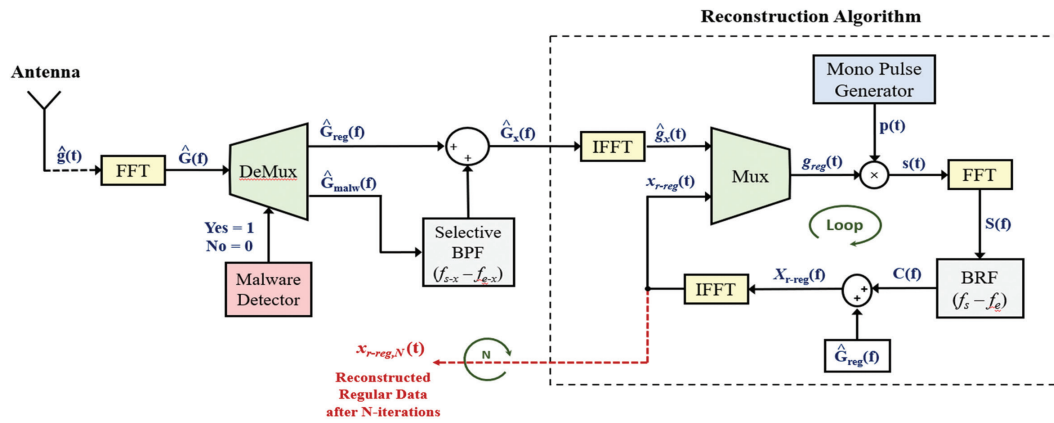


Fig. 14. The proposed reconstruction algorithm in the RFID reader

Table 4. Comparison of Signal Reconstruction and Noise Reduction Techniques for Various Noisy Systems and Applications

Ref. (Year)	Signal Type	System/ Application	Reconstruction Algorithm	Noise Type	Noise Reduction Technique	Performance Metrics & Result
[23] (2020)	Stream of pulses	IoT and mobile WSN	Sparse pulse representation for signal reconstruction	AWGN	Using denoising autoencoders	- Proposed method: LPR = 0.01, SPR = 0.02 - Conventional method: LPR = 0.05, SPR = 0.1
[24] (2021)	OFDM	IBFD communication in OFDM-based wireless systems	DSIC techniques, for estimating the SI channel and reconstruction of the SI signal	CP noise	CPNR	- Improved total suppression by 6 db. - Improved EVM by 5%
[25] (2022)	ECG pulses	Portable heartbeat detection system	DWT	AWGN	A combination of DWT and SVD	Proposed method average: MSE = 0.002, PSNR = 60.5 dB
[26] (2022)	Stream of pulses	X-ray single-particle imaging	A neural network pipeline for restoring diffraction intensities.	Poisson noise	Poisson noise reduction	An improvement in the MSE of roughly two orders of magnitude.
[27] (2021)	dMRI signal	MRI based systems	KPCA	AWGN	KPCA denoising	Better SNR improvements up to 2.7x
[28] (2022)	OFDM	UWOC systems	CS- based channel estimation	AWGN	Setting a noise threshold to remove useless channel taps.	Increased NMSE by 67% and 97% than and 97% than DFT and LS algorithms respectively
Proposed Method (2023)	Stream of pulses	Baseband and speech signals	Proposed Baseband signal reconstruction algorithm	AWGN	NRF	An improved average NMSE by 47.7% after 100 iterations

Figure 13 depicts the implementation of the proposed noise reduction scheme for reconstructing baseband signals from random streams of pulses with different data rates (2, 4, and 6 Mbps), transmitted window bandwidth of 0.3–8.3 MHz, and varying numbers of iterations, at an SNR of 10 dB. The received stream signals with noise for the different data rates are denoted by $g_{n-SX}(t)$, while $x_{r-SX}(t)$ and $x_{sX}(t)$ represent the reconstructed and original generated signals, respectively, which are compared to each other in each subfigure. The figure clearly illustrates the successful application of our proposed noise rejection scheme to reconstructing different data streams. The NMSE initially decreases as the number of iterations (N) increases but reaches a certain value beyond which it diverges and results in a worse NMSE value. For example, signal reconstruction for the 2 Mbps stream is perfect when the algorithm is run for up to 45 iterations, but beyond this point, the NMSE increases, leading to bad signal reconstruction status. Similarly, the optimal number of iterations for the 4 Mbps and 6 Mbps streams are 50 and 103, respectively. It is worth noting that our noise reduction scheme using the NRF has limitations in that it cannot be applied for any number of iterations, as it converges to a certain number of iterations and then diverges.

Table 4 shows a comparison between different signal reconstruction and noise reduction schemes and their state-of-the-art applications. Their evaluation metrics are also compared in terms of lost pulse ratio (LPR), spurious pulse ratio (SPR), mean squared error (MSE), NMSE, and peak signal-to-noise ratio (PSNR). In [23], the authors propose a denoising autoencoder-based sparse pulse representation method for internet of things (IoT) and mobile wireless sensor network (WSN) systems where the proposed method achieved lower values of LPR and SPR compared to the conventional method. In [24], the authors use different digital self-interference cancellation (DSIC) techniques for self-interference (SI) channel estimation and reconstruction in orthogonal frequency division multiplexing (OFDM)-based wireless systems. The proposed DSIC technique resulted in a 6 dB improvement in total suppression and a 5% improvement in error vector magnitude (EVM). In [25], the authors combine discrete wavelet transform (DWT) and singular value decomposition (SVD) for ECG pulse denoising in a portable heartbeat detection system. The proposed method achieved an improved MSE of 0.002 and a PSNR of 60.5 dB.

In [26], a neural network pipeline for restoring diffraction intensities in X-ray single-particle imaging systems affected by Poisson noise is proposed, and this method resulted in an improvement in the MSE of roughly two orders of magnitude. The authors in [27] apply kernel principal component analysis (KPCA) denoising to diffusion magnetic resonance imaging (dMRI) signals in MRI-based systems. The proposed KPCA approach led to better SNR improvements of up to 2.7x. In [28], the authors employ a compressed sensing (CS)-based

channel estimation technique and set a noise threshold to remove useless channel taps in underwater optical wireless communication (UWOC) systems affected by AWGN. The proposed approach outperformed discrete fourier transform (DFT) and least squares (LS) algorithms in terms of NMSE by 67% and 97%, respectively. Finally, the proposed approach applies a baseband signal reconstruction algorithm to AWGN in stream of pulses, baseband, and speech signals, achieving an improved average NMSE by 47.7% after 100 iterations.

3.3. PROPOSED APPLICATIONS

In [22], the proposed baseband signal reconstruction algorithm was applied to different signal types such as pulse, triangular, composite, analog, stream of pulses, and speech signals, and the algorithm showed its success in reconstructing most of their signals' energy and thus reconstructing them. The proposed scheme for baseband signal reconstruction can be used in various applications to reconstruct signals. For example, in UWB communications, it can reconstruct short-width pulses, while in radio frequency identification (RFID) systems, it can extract data transmitted by tags. Additionally, it can optimize bandwidth in mobile communication systems and recover pulses in radar systems. It can also optimize bandwidth in speech and audio processing applications.

In [29], a proposed real-world application in ultra-high frequency (UHF) RFID systems in which the proposed algorithm is used to reconstruct the bit stream of regular data in malware-free and malware-injected scenarios, as shown in Fig. 14. In a malware-free scenario, the received signal $g(t)$ is fed to a fast Fourier transform (FFT) block and then identified as malware-free or malware-injected data by the malware detector block. If it is malware-free, the data is passed directly to the reconstruction algorithm to recover the original regular data, but if it is malicious, it extracts the missing spectrum of the original regular data from the malicious data spectrum and then adds it to the received regular data spectrum, and then the proposed reconstruction algorithm is applied to recover the regular data more quickly.

4. CONCLUSION AND FUTURE WORK

This paper proposes a noise reduction scheme to minimize the effect of noise on a proposed algorithm used to reconstruct the baseband signals. The algorithm is based on sending small-transmitted window that carries a portion of the original signal's energy, and then reconstructing the signal again, thus optimizing the used channel bandwidth. Although the algorithm converges in noise-free channels, it diverges when noise is present, leading to bad reconstruction status in each iteration of the algorithm. At the receiver, the noise can be minimized by using a traditional way by compromising between three parameters, the transmitted window bandwidth, the number of iterations and SNR, but this is not an effective solution. So,

a noise reduction scheme based on adaptive NRF is proposed which is a function of the system's starting and ending frequencies. When tested on a baseband pulse and a stream of pulses in the presence of AWGN, the suggested scheme showed good reconstruction performance in reducing the noise effect. The performance was evaluated in terms of NMSE compared to free noise channels. It was applied to different streams of pulses with various data rates (2, 4, 6, and 8 Mbps) and demonstrated good reconstruction performance in terms of NMSE, but only up to a certain number of iterations, after which the proposed scheme diverged. This algorithm may find applications in future sixth generation (6G) wireless networks, UWB communications, radar systems, RFID-based systems, and vehicle-to-vehicle communications (V2V). Future research can further study this algorithm in these cases.

5. REFERENCES

- [1] M. M. Richter, S. Paul, V. Kėpuska, M. Silaghi. "Digital Signal Representation", In *Signal Processing and Machine Learning with Applications*, Springer, 2022, pp. 3-38.
- [2] Y. Yamamoto, M. Nagahara, P. P. Khargonekar, "A brief overview of signal reconstruction via sampled-data H_∞ optimization", *Applied and Computational Mathematics*, Vol. 11, No. 1, 2012, pp. 3-18.
- [3] L. Guo, C. W. Kok, H. C. So, W. S. Tam, "Fast and L_2 -optimal recovery for periodic nonuniformly sampled bandlimited signal", *Signal Processing*, Vol. 180, 2021, p. 107856.
- [4] Z.-C. Zhang, "Nonuniform reconstruction of periodic bandlimited signals without sampling points' number restriction", *Optik*, Vol. 207, 2020, p. 163798.
- [5] H. Zhao, R. Wang, D. Song, "Recovery of Bandlimited Signals in Linear Canonical Transform Domain from Noisy Samples", *Circuits, Systems, and Signal Processing*, Vol. 33, No. 6, Jan. 2014, pp. 1997-2008.
- [6] V. Kishore, S. Mukherjee, C. S. Seelamantula, "Phase Sense—Signal Reconstruction from Phase-Only Measurements via Quadratic Programming", *Proceedings of the International Conference on Signal Processing and Communications*, Bangalore, India, 19-24 July 2020, pp. 1-5.
- [7] R. Balan, B. G. Bodmann, P. G. Casazza, D. Edidin, "Fast algorithms for signal reconstruction without phase", *SPIE Proceedings*, Vol. 6701, 2007.
- [8] S.-S. Zuo, M. Xing, X.-G. Xia, G.-C. Sun, "Improved Signal Reconstruction Algorithm for Multichannel SAR Based on the Doppler Spectrum Estimation", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 4, 2017, pp. 1425-1442.
- [9] J. Chen, S. Sun, N. Bao, Z. Zhu, L. -B. Zhang, "Improved Reconstruction for CS-Based ECG Acquisition in Internet of Medical Things", *IEEE Sensors Journal*, Vol. 21, No. 22, 2021, pp. 25222-25233.
- [10] Z. Zhang et al. "Electrocardiogram Reconstruction Based on Compressed Sensing", *IEEE Access*, Vol. 7, 2019, pp. 37228-37237.
- [11] M. Al Disi et al. "ECG Signal Reconstruction on the IoT-Gateway and Efficacy of Compressive Sensing Under Real-Time Constraints", *IEEE Access*, Vol. 6, 2018, pp. 69130-69140.
- [12] Z. Luo, J. Liang, J. Ren, "Deep Learning Based Compressive Sensing for UWB Signal Reconstruction", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, 2022, pp. 1-10.
- [13] K. Abratkiewicz, P. Samczyński, "A Block Method Using the Chirp Rate Estimation for NLFM Radar Pulse Reconstruction", *Sensors*, Vol. 19, No. 22, Nov. 2019, p. 5015.
- [14] G. Baechler, A. Scholefield, L. Baboulaz, M. Vetterli, "Sampling and Exact Reconstruction of Pulses with Variable Width", *IEEE Transactions on Signal Processing*, Vol. 65, No. 10, 2017, pp. 2629-2644.
- [15] T. Zahavy, A. Dikopoltsev, O. Cohen, S. Mannor, M. Segev, "Deep Learning Reconstruction of Ultrashort Pulses", *Optica*, Vol. 5, No. 5, 2018, pp. 666-673.
- [16] G. Baechler, A. Scholefield, L. Baboulaz, M. Vetterli, "Sampling and Exact Reconstruction of Pulses with Variable Width", *IEEE Transactions on Signal Processing*, Vol. 65, No. 10, May 2017, pp. 2629-2644.
- [17] A. G. J. Besson, "Imaging from Echoes: On Inverse Problems in Ultrasound", *École Polytechnique Fédérale de Lausanne, Switzerland*, PhD Thesis, 2019.
- [18] S. Xu, S. Tao, Y. Chai, X. Yang, Y. He, "The extrapolation of bandlimited signals in the offset linear canonical transform domain", *Optik*, Vol. 180, 2019, pp. 626-634.

- [19] J. Weng, "A new one-step band-limited extrapolation procedure using empirical orthogonal functions", *Journal of Electronics*, Vol. 23, No. 5, 2006, pp. 777-780.
- [20] M. Milman, "Extrapolation and optimal decompositions: with applications to analysis", Springer-Verlag, 1994.
- [21] A. F. Ashour, A. Khalaf, A. Hussein, H. Hamed, A. Ramadan, "A New Algorithm for Baseband Pulse Transmission over Band-Limited Channels for Wireless Automotive Communications", *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 4, 2020, pp. 5222-5228.
- [22] A. F. Ashour, A. Khalaf, A. Hussein, H. Hamed, A. Ramadan, "A Proposed Signal Reconstruction Algorithm over Bandlimited Channels for Wireless Communications", *Advances in Electrical and Computer Engineering*, Vol. 23, No. 1, 2023, pp. 19-32.
- [23] X. Li, Z. Liu, Z. Huang. "Deinterleaving of pulse streams with denoising autoencoders", *IEEE transactions on aerospace and electronic systems*, Vol. 56, No. 6, 2020, pp. 4767-4778.
- [24] H. Ayar, O. Gurbuz, "Cyclic prefix noise reduction for digital self interference cancellation in OFDM-based in-band full-duplex wireless systems", *IEEE Transactions on Wireless Communications* 20, No. 9, 2021, pp. 6224-6238.
- [25] G. Huang, Z. Yang, W. Lu, H. Peng, J. Wang, "Sub-Nyquist sampling of ECG signals based on the extension of variable pulsewidth model", *IEEE Transactions on Instrumentation and Measurement*, Vol. 71, 2022, pp. 1-14.
- [26] A. Bellisario, F. R. N. C. Maia, T. Ekeberg. "Noise reduction and mask removal neural network for X-ray single-particle imaging", *Journal of Applied Crystallography*, Vol. 55, No. 1, 2022, pp. 122-132.
- [27] G. Ramos-Llordén et al. "SNR-enhanced diffusion MRI with structure-preserving low-rank denoising in reproducing kernel Hilbert spaces", *Magnetic Resonance in Medicine* 86, No. 3, 2021, pp. 1614-1632.
- [28] X. Liu, J. Hu, K. Zhang, X. Tang, Y. Dong, "On Channel Estimation Based on Compressed Sensing for OFDM UWOC Systems", *Proceedings of the Asia Communications and Photonics Conference*, Shenzhen, China, 5-8 November 2022, pp. 1039-1042.
- [29] A. F. Ashour, C. Condie, C. Pocock, S. C. Chiu, A. Chrysler, M. M. Fouda, "A Spectrum Injection-Based Approach for Reconstructing Regular Data from Malware-Injected Data in UHF RFID Systems", *Proceedings of the IEEE Global Communications Conference*, 2023. (in press)

Lucy Richardson and Mean Modified Wiener Filter for Construction of Super-Resolution Image

Original Scientific Paper

Pravin Balaso Chopade

Department of Electronics & Telecommunications,
M.E.S College of Engineering, Pune, India
pbchopade@mescoepune.org

Prabhakar N. Kota

Department of Electronics & Telecommunications,
M.E.S College of Engineering, Pune, India
Prabhukota1@gmail.com

Bhagvat D. Jadhav

Department of Electronics & Telecommunications,
JSPM's Rajarshi Shahu College of Engineering,
Pune, India
Bdjadhav_entc@jspmrscoe.edu.in

Pravin Marotrao Ghate

Department of Electronics & Telecommunications,
JSPM's Rajarshi Shahu College of Engineering,
Pune, India
pmghate_entc@jspmrscoe.edu.in

Shankar Dattatray Chavan

Department of Electronics & Telecommunications,
Dr. D. Y. Patil Institute of Technology, Pimpri, India
sdchavan73@gmail.com

Abstract – The ultimate goal of the Super-Resolution (SR) technique is to generate the High-Resolution (HR) image by combining the corresponding images with Low-Resolution (LR), which is utilized for different applications such as surveillance, remote sensing, medical diagnosis, etc. The original HR image may be corrupted due to various causes such as warping, blurring, and noise addition. SR image reconstruction methods are frequently plagued by obtrusive restorative artifacts such as noise, stair casing effect, and blurring. Thus, striking a balance between smoothness and edge retention is never easy. By enhancing the visual information and autonomous machine perception, this work presented research to improve the effectiveness of SR image reconstruction. The reference image is obtained from DIV2K and BSD 100 dataset, these reference LR image is converted as composed LR image using the proposed Lucy Richardson and Modified Mean Wiener (LR-MMWF) Filters. The possessed LR image is provided as input for the stage of bicubic interpolation. Afterward, the initial HR image is obtained as output from the interpolation stage which is given as input for the SR model consisting of fidelity term to decrease residual between the projected HR image and detected LR image. At last, a model based on Bilateral Total Variation (BTV) prior is utilized to improve the stability of the HR image by refining the quality of the image. The results obtained from the performance analysis show that the proposed LR-MMW filter attained better PSNR and Structural Similarity (SSIM) than the existing filters. The results obtained from the experiments show that the proposed LR-MMW filter achieved better performance and provides a higher PSNR value of 31.65dB whereas the Filter-Net and 1D,2D CNN filter achieved PSNR values of 28.95dB and 31.63dB respectively.

Keywords: bilateral total variation, fidelity term, lucy richardson filter, modified mean wiener filter, super-resolution

1. INTRODUCTION

Technology in both software and hardware has advanced significantly during the past two decades. Industrial sectors have used modern technology to its fullest potential to produce electronic gadgets like computers, mobiles, Personal Digital Assistants (PDA), and countless gadgets at low prices [1]. To produce images with high quality, camera sensor manufacturing techniques have also advanced significantly. Digital sensors are designed to take a Low-Resolution (LR) image as an input and produce a High-Resolution (HR) image. The resulting

HR image is anticipated to have sufficient edge information artifacts [2, 3]. Super-Resolution (SR) is a technique for generating high-quality images or frames from their low-quality counterparts utilizing digital image processing techniques. Applications of SR images are currently being used in academics and industry [4, 5]. SR is a developing field with a vast number of applications in electronic imaging, including forensics, surveillance, satellite imaging, and more [6]. The SR techniques utilize edge information to moderate the distorted problems in the tasks for obtaining super-resolution images. The SR techniques are categorized as the extraction of edges

and conversion of super-resolution images [7]. The SR technique is utilized to perform non-linear mappings among LR-HR images based on conventional super-resolution methods [8].

In real-time applications of SR imaging systems, numerous factors are responsible to lower the quality of the image. This is due to physical limitations, less count of image sensors, and a lower spatial rate for image sampling [9]. The SR technique is applied to the normal image sources to increase the image quality by adjusting contrast, saturation, etc [10]. The conversion of LR image to HR image is performed using super-resolution techniques such as Single Image Super-Resolution (SISR) and Multi-Image Super-Resolution (MISR). Among two techniques, SISR is widely used in the process of HR construction because it focuses on an individual LR image to provide a better result [11]. But, MISR can fill up the unavailable information by collecting the related data from other image sources. Conventionally, both techniques are utilized according to the user's needs [12, 13]. Many SR methods get the input of LR by relating it with degradation methods based on ground-truth images [14]. In some cases, the process of super-resolution is utilized to extract the information from LR images and achieve better results of spatial preservation [15,16]. In this paper, the quality of the reference LR image is enhanced using the proposed Modified Median Wiener Filter to overcome the issues related to image reconstruction process.

The main contribution of this research is listed as follows:

- (i) The MMW filter cannot preserve the sharp features of the images (E.g. Lines). To overcome these drawbacks, the hybrid-MMW filter is proposed, which can preserve the sharp edges and eliminate the noise present in the image. Moreover, it smoothens the image by decreasing the intensity variations in the neighboring pixels and removing the blurred parts of the image.
- (ii) The quality of the initial HR image is enhanced from the interpolation stage which is given as the input for the SR model.
- (iii) The fidelity term is created to decrease the residual between the projected HR image and the detected LR image.

The remaining paper is organized as follows, Section 2 represents the related works of the paper. The proposed method is discussed in Section.3. The results and analysis are presented in Section 4. Finally, Section 5 represents the overall conclusion of the paper.

2. RELATED WORKS

This section provides related works on various filtering methods used to obtain SR images.

Park et al. [17] introduced a deep learning model consisting of 1D and 2D filters for SR of images. The 1D layer for extraction of features and 2D restoration layers of

the image made up the suggested model. Using 1D filters, the first layer for extraction of features was created to separate parallel and perpendicular high-frequency signals. The second HR image-restoration layers were created to extract high-frequency signals by 2D filters. With the least amount of visual loss, the deep learning model was used to decrease the complexity of super-resolution methods. However, the model needed complex calculations and used more memory.

Luo et al. [18] introduced a lightweight Super-resolution model known as LatticeNet which utilized lattice blocks connected in series and backward feature fusion. The model combined multiple residual blocks by combinational co-efficient and this combinational co-efficient is present in any super-resolution block using residual blocks as a basic block. Moreover, backward fusion was utilized to combine information related to various fields. The structure of the lattice block consisted of two residual blocks in a linear combination that helped to increase the chance to attain a dominant network. However, the model can't be utilized in real-time applications and required more storage memory.

Esmaeilzahi et al. [19] introduced a Multiple spatial Range and Resolution level feature by generating a deep recursive Network (MuRNet) which initialized from the bicubic interpolated versions of images with low resolution and created high-quality super-resolved images. The introduced method combined multi-spatial range and level of resolution to provide enhanced maps of features in the framework of a recursive network. The MuRNet provided less count in the multiplied operations and improved the performance of the network. But the image was handled poorly, as it created an issue in the vanishing of gradient and lowered the overall performance of the super-resolution image.

Chen et al. [20] introduced a super-resolution image reconstruction technique based on an attention mechanism along with the feature map to enable images with low resolution to be super-resolute images. The reconstruction model contained three blocks for the extraction of features, the extraction of images, and a module to perform the reconstruction. The introduced technique increased the super-resolution effect of the image and improved the evaluation of quantitative objectives. However, the introduced technique performed slower in the reconstruction process due to its less capability.

Li et al. [21] introduced a deep adaptive information Filtering Network (Filter Net) to perform fast and accurate image Super-resolution. The introduced Filter Net could filter the information with low frequency from the adaptive features. The Filter Net contained dilated residual group with multi units of dilated residuals which improved the receptive field and exploited the information of low-resolution input. The Filter Net utilized an adaptive information fusion structure to construct weighted connections to increase the pixel-fitting capacity in the network. However, the consumption time was high and more memory space was occupied.

Gao and Zhou [22] introduced a Very Lightweight and Efficient Single Image Super Resolution (VLESR) method to balance the pixels of the image to get SR images. The VLESR technique utilized a Light Weight Residual Concatenation Block (LRCB) which propagates and fuses the local features of the image. Additionally, the Multi-way attention Block has the capability to integrate the features and enhance the quality of SR restoration of the image. However, the introduced VLESR did not contain any channel Concatenation and 1×1 convolution. If it's present it probably enhances the image accuracy.

He et al. [23] introduced an Orientation Aware feature extraction Model (OAM) that comprised a mixture of 1D and 2D kernels. In every OAM, the channel attention mechanism was employed to perform specific scene selections. The high-quality SISR combined both high-level and low-level features through progressive fusion Orientation method, which led to a concise execution of the SISR task.

3. LUCY RICHARDSON AND MEAN MODIFIED WIENER FILTER FOR CONSTRUCTION OF SUPER-RESOLUTION IMAGE

The presence of noisy and blurring effects are the general issues in SR images. A Bayesian framework is taken into consideration for implanting prior data into HR images to rectify this issue. The estimated vector is taken to be Gaussian because it is known that the HR image contains white Gaussian noise. MAP framework is used to solve the minimization problem as a result of the Bayes rule's procedure of creating the HR image. The first HR image is estimated via the conventional SR methodologies using bicubic or bilinear interpolation techniques on the reference LR image. However, LR image with low quality is obtained which leads to poor performance in the estimation of HR image. To overcome this issue and to obtain a composed LR image, the Median Modified Wiener Filter (LR-MMWF) is used. The composed LR image is fed into the bicubic interpolation to obtain the initial HR image. Then, the initial HR image is processed using the super-resolution model which consists of fidelity terms and regularization terms to provide the final HR image. The overall process involved in obtaining the SR image is represented in Fig. 1.

3.1. STEPS TO OBTAIN SUPER-RESOLUTION IMAGE

The process involved in the conversion of reference LR image into a super-resolution image using LR-MMWF involves various steps mentioned as follows:

- (i) A series of low-resolution images are generated from the DIV2K dataset and BSD 100 which is considered as the input for the conversion process.
- (ii) Assumption is made for one low-resolution image which is considered as reference LR image. The reference LR image is utilized to initialize the process to obtain a high-resolution image.
- (iii) The angle of rotation, vertical and horizontal shifts

that exist in the reference LR image is computed to improve the warp matrix of the image.

- (iv) The process of conversion is employed in reference LR image to obtain the possessed LR image represented in Fig. 1. The composed LR image is given as the input for the stage of bicubic interpolation.
- (v) The bicubic interpolation is applied to the possessed LR image to compute a high-resolution image at the initial stage. This initial resolution image is provided as an input for the stage of reconstruction of the image which is explained in section 3.1
- (vi) The initial HR image is given as the input for the super-resolution model for image reconstruction. There the model computes the Fidelity term and regularization term using equation (6) and produces the final HR image as a super-resolution image.

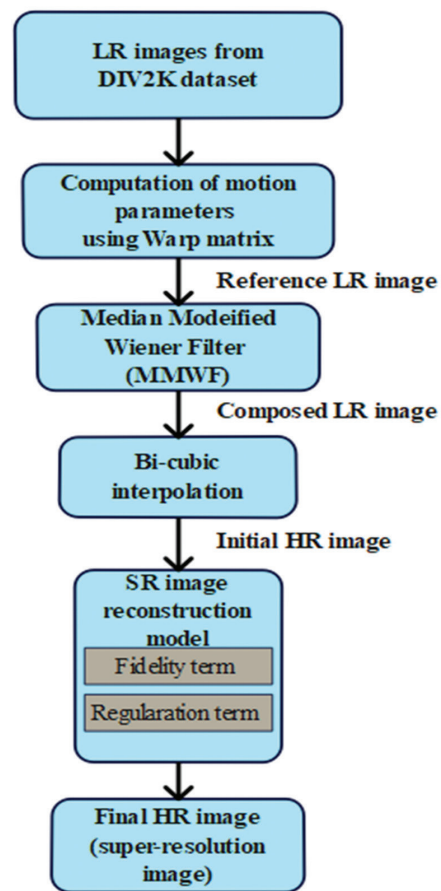


Fig. 1. Overall Process to Obtaining SR image

3.2. DATASET

The work is processed using two image datasets such as BSD100 [24] and DIV2K [25], and this section provides a brief description of these datasets. The Berkeley Segmentation Dataset 100 (BSD-100) consists of 100 images related to nature. A well-known image super-resolution dataset called DIV2K has 1,000 images of various scenarios, 800 of which are for training, 100 for validation, and 100 for testing. To promote research on image super-resolution with more realistic deterioration, it was gathered for the NTIRE2017 and NTIRE2018 Super-Resolution Challenges.

Low-resolution image with various sorts of degradations is included in this dataset. In addition to the typical bicubic downsampling, other degradations are taken into account while creating low-resolution images for the different challenge tracks. The three first images of the Div2K dataset and their corresponding gray and binary image are represented in Fig. 2 and some sample images from the BSD-100 dataset are provided in Fig. 3.



Fig. 2. The three first images of the Div2K dataset and their corresponding gray and binary image



Fig. 3. Sample images from the BSD-100 dataset

3.3. FRAMEWORK FOR SUPER-RESOLUTION IMAGES

The images obtained from the dataset are provided for the Bayesian framework to compute the possibility of the problem of uncertainty based on available knowledge by conjoining various priors into a mathematical model that is capable of statistical implication. The Bayesian framework is used to estimate the likelihood of uncertainty problems depending on currently accessible knowledge. Hence, the uncertainty inference problem occurs in SR technology for evaluating projected HR images from a series of detected LR images. In other words, the observed LR image provides proof for the deduction of the HR image, and the reconstructed SR solution constrains the regularization term present in the image. SR

reconstruction is equal to estimating the HR picture using provided LR images in the Bayesian framework. The Maximum Posteriori (MAP) technique, combines the image's prior constraints and produces outcomes by maximizing the probability cost function, which is a well-preferred one for the Bayesian framework. Additionally, it is known for its adaptability in the estimation of joint parameters and the preservation of edges. In general, rather than using the given parameters, Bayesian estimation is used to evaluate the probability distribution for unknown parameters. The function of the MAP estimator is mathematically defined using the equation (1)

$$\hat{X}_{MAP} = \arg \max_X \prod_{k=1}^M P(Y_k|X). P(X) \quad (1)$$

Where the probability conditions of low-resolution images (Y_k) from the HR image (X) are represented as $P(Y_k|X)$ and the probability of the HR image at the prior condition is represented as $P(X)$.

The significant structure of SR consists of fidelity term and term of regularization. The MAP estimator is applied for reduction of residues among the HR image and LR image. Regularization is performed to improve the stability of the HR image. The mathematical form of the super-resolution framework is represented in equation (2).

$$\hat{X} = \arg \min_X \sum_{k=1}^N \rho(DHF_k X - Y_k) + \lambda R(X) \quad (2)$$

Where the term data fidelity is represented as $\rho(DHF_k X - Y_k)$, the term of regularization is denoted as $R(X)$, and the parameter of regularization is represented as λ .

3.4. CONVERSION OF REFERENCE LR IMAGE TO COMPOSED LR IMAGE

The Bayesian framework provides a reference LR image as an input for the conversion of the reference LR image into a composed LR image. The median filter and the proposed LR-MMWF are utilized to convert the reference LR image to a composed LR image. The initial high-resolution image is assessed via the method of bicubic interpolation by providing a possessed LR image as input. The median filter smoothens the image by decreasing the variation intensity among the nearby pixels and the Lucy Richardson filter is used in removing the blurred part of the image. To reduce the residual between the projected HR image and the detected LR image, the L_2 is used for the data-fidelity term. Finally, the minimized operation is limited to the stabilized condition of the created HR image using the BTW prior model. The process of converting a reference LR image into a combined LR image is represented in Fig. 4.

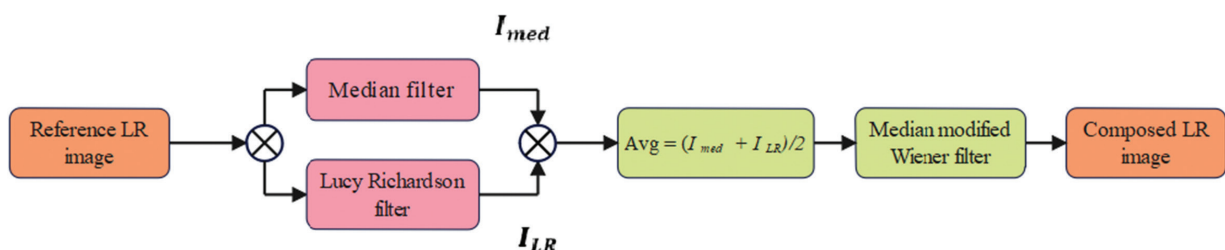


Fig. 4. Conversion of reference LR image to composed LR image

3.4.1. Hybrid of Lucy Richardson and Median Modified Wiener (LR-MMW) Filter

The reference image obtained from the Bayesian framework is processed using an LR-MMW filter to produce a composed LR image where the inhibited image quality is obtained. The widely utilized de-noising techniques in image processing are performed using linear filters (spatial filters). The spatial filter is based on the kernel's size which is generally referred to as a mask, it uses the intensity of the pixel to calculate the new pixel value. The median filters are less sensitive to extreme values, generally known as outliers. Moreover, the median filter efficiently neglects extreme values without affecting the image's sharpness. This advantage of the median filter is embedded with the advantage of the Wiener filter in producing edge-preserved images to make the MMW filter. Secondly, the Lucy Richardson filter is used in the process of de-blurring the image to provide an image with high resolution. The goal of the LR-MMWF strategy relies on increasing the quality of the image by de-noising the region of the background of a degraded image utilizing a median filter. This method is used to minimize the noise distribution in degraded images. Additionally, this method largely uses the Wiener filter to maintain the edge signal of the image. The LR-MMWF method is based on the Wiener filter and Lucy Richardson filter to reduce noise in the deteriorated image by replacing the value of pixels in the mask matrix along with median values. The average computed value in the Wiener filter gets interchanged by the value from the median filter. It is computed using equation (3) as follows:

$$b_{MMWF} = \bar{\mu} + \frac{\sigma^2 - v^2}{\sigma^2} \cdot (a - \bar{\mu}) \quad (3)$$

Where $\bar{\mu}$ is the median value, the variance in the Gaussian noise of the image is represented as σ^2 , and the variance of noise in the matrix of the Wiener filter is denoted as v^2 . The area of every individual pixel is represented as a .

The benefit of employing the LR-MMWF technique is that the edge signal is better retained when compared to the usage of the median and Wiener filter techniques. This helps to improvise the quality of the deteriorated image using the drop-off effect. Regarding the de-noising effect, the LR-MMWF technique can significantly outperform traditional filters. In addition, it can simultaneously keep the edge signal and eliminate the background noise signal in the image.

3.5. RECONSTRUCTION OF IMAGE

The composed LR image obtained from a combination of LR-MMW filters is reconstructed to provide a super-resolution image from the HR image. Here the data fidelity term is represented as L_p and the regularization term is discussed using the model of BTV (Bilateral Total Variation). The data fidelity L_p is defined for the high-resolution image using equation (4).

$$F(X) = \arg \min_X \sum_{k=1}^N (DHF_k X - Y_k) \quad (4)$$

where the motion and blur matrices are represented as D and H , and the movement of the matrices is denoted as F .

The BTV model is a combination of a bilateral filter with total variation, and it utilizes more neighbors to measure the gradient level for the provided pixels. Moreover, the sharp edges of the image are preserved using BTV model as it is low in cost and easy for implementation. The formula of BTV is mathematically represented in equation (5).

$$R_{BTV}(X) = \sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|} X - S_x^l S_y^m X_1 \quad (5)$$

Where the shifted HR image in the horizontal direction for l pixels is denoted as S_x^l and the shifted HR image in the vertical direction for m pixels is denoted as S_y^m (where $l+m \geq 0$). The scaling weight is represented as α which lies in the range of $0 < \alpha < 1$ and P is known as the controlled parameter.

The problem can be addressed in an infinite number of ways and under various circumstances. Additionally, minute levels of noise present in the measurements, cause significant agitations in the final solution, making it unstable. Therefore to arrive at a stable solution, artifacts are removed from the final solution, the convergence rate is improved, and the regularization method is used in SR image reconstruction. There are many regularization techniques, and one of them has is used to produce HR images with sharp edges and implementation simplicity as a primary advantage. In the picture reconstruction step, the regularization term is used to make up for missing data using some prior knowledge that is used to reduce loss. The regularization of BTV is formulated using equation (6)

$$X = \arg \min_X \left[\sum_{k=1}^N (DHF_k X - Y_k) + \lambda \sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|} X - S_x^l S_y^m X_1 \right] \quad (6)$$

The steepest descent method is utilized to obtain a fast convergence rate for the true HR image and it is assessed with less count of iterations. The iteration using the steepest descent method is computed by the formula in equation (7).

$$X_{n+1} = X_n - \beta \left[\sum_{k=1}^N F_k^T H^T D^T \text{sign}(DHF_k X_n - Y_k) + \lambda \sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|} [I - S_y^{-m} S_x^{-l}] \text{sign}(X_n S_x^l S_y^m X_n) \right] \quad (7)$$

Where the scalar unit which regulates size of the step in the gradient direction is represented as β and the parameter utilized for regularization is denoted as λ . The terms S_y^{-m} and S_x^{-l} represent the transposed matrices of S_y^m and S_x^l correspondingly.

4. RESULTS AND DISCUSSION

This section provides results obtained from the proposed LR-MMW filter for various metrics to evaluate image quality. The design and simulation of the LR-MMW filter can be performed using MATLAB software R2020a and system specifications with an Intel Core i7 processor with 8 GB RAM and a 64-bit Windows 10 operating system. The LR-MMW filter is utilized to provide

super-resolution images from the considered datasets such as DIV2K and BSD100 dataset. The performance of the LR-MMW filter is evaluated for parameters such as Peak-Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to compute the quality of the image. The mentioned evaluation metrics (PSNR and SSIM) can be computed using equation (9) and equation (13) as follows

PSNR

The PSNR is one of the general methods to evaluate the image quality. It is usually described using Mean Square Error (MSE). The MSE for two $W \times H$ images X and Y is computed using the formula given in equation (8).

$$MSE = \frac{1}{WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [X(i, j) - Y(i, j)]^2 \quad (8)$$

Hence, the PSNR is represented by equation (9)

$$PSNR = 10 \log \left(\frac{X_{MAX}^2}{MSE} \right) = 20 \log \left(\frac{X_{MAX}}{\sqrt{MSE}} \right) \quad (9)$$

Where the largest pixel value of the image is denoted as X_{MAX} .

SSIM

The structural similarity of the image is provided by the intensity $L(x, y)$ of the image, contrast of the image $C(x, y)$ and structure of the image $S(x, y)$.



Fig. 5. Input LR image from DIV2K dataset

The value of $L(x, y)$, $C(x, y)$ and $S(x, y)$ is computed using the formula provided in the equation (10-12) as follows:

$$L(x, y) = (2\mu_x\mu_y + C_1) / (\mu_x^2 + \mu_y^2 + C_1) \quad (10)$$

$$C(x, y) = (2\sigma_x\sigma_y + C_2) / (\sigma_x^2 + \sigma_y^2 + C_2) \quad (11)$$

$$S(x, y) = (\sigma_{xy} + C_3) / (\sigma_x\sigma_y + C_3) \quad (12)$$

Where the reconstructed HR image is denoted as x and the original HR image is denoted as y . The mean and variance of the image are denoted as μ and σ respectively. The constants are denoted as C_1, C_2 , and C_3 respectively.

Hence, the SSIM is defined using the equation (13) as follows:

$$SSIM(x, y) = L(x, y) \cdot C(x, y) \cdot S(x, y) \quad (13)$$

Where $L(x, y)$ is denoted as intensity of the image, $C(x, y)$ is denoted as contrast of the image, and the structure of the image is denoted as $S(x, y)$.

The input LR image and output SR image using Lucy Richardson and Mean Modified Wiener Filter are represented in Fig. 5 and Fig. 6 respectively. The size of the input LR image is 481×321 and the size of the output image is 850×641



Fig. 6. Output SR image from the DIV2K dataset

4.1. PERFORMANCE ANALYSIS

The performance of the proposed LR-MMW filter for the DIV2K and BSD100 datasets is described in this section. The performance is evaluated based on parameters such as Peak-Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The performance is evaluated for the Kalman filter (KF), Wiener Filter (WF), Mean Modified Wiener Filter (MMWF), and the proposed LR-MMW filter. Table.1 represents the performance of the proposed MMWF filter for the DIV2K dataset.

Table 1. Performance of various filters for DIV2K dataset

Dataset	Filters	PSNR (dB)	SSIM
Diverse2K (DIV2K) dataset	Wiener Filter	21.43	0.784
	Kalman Filter	24.56	0.813
	MMW Filter	29.62	0.886
	LR-MMW Filter	31.69	0.92

As per Table 1, the performance of various filters to improve image quality is discussed by computing the parameters such as PSNR and SSIM.

The results obtained from Table 1 show that the proposed LR-MMW filter attained high PSNR and SSIM of 31.54 dB and 0.912 and provides an effective SR image from the DIV2K dataset. This better result is due to the presence of an image reconstruction model which consists of (L_2 Norm) and (BTV prior). The L_2 Norm lessens the residual in the predicted HR image and BTV improves the stable state of generated HR image. The graphical comparison of PSNR value for various filters including the proposed LR-MMW filter is represented in Fig. 7:

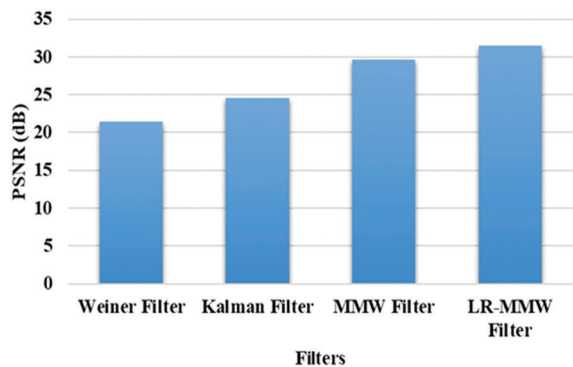


Fig. 7. Graphical representation of PSNR value for DIV2K dataset

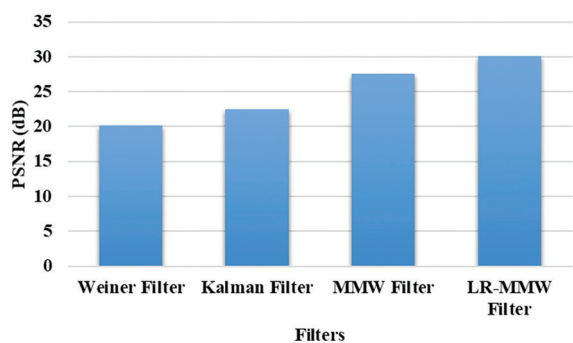


Fig. 8. Graphical representation of PSNR value for BDS 100 dataset

In Table 2 the performance of various filters to improve the image quality is discussed by computing the parameters such as PSNR and SSIM. The results obtained from Table 2 show that the proposed LR-MMW filter attained high PSNR and SSIM of 31.65 dB and 0.912 respectively for the BDS100 dataset and provides an effective super-resolution image. The graphical comparison of PSNR value for various filters including the proposed LR-MMW filter is represented in Fig. 8.

Table 2. Performance of various filters for the BDS 100 dataset

Dataset	Filters	PSNR (dB)	SSIM
Berkeley Segmentation Dataset (BSD-100)	Wiener Filter	20.12	0.762
	Kalman Filter	22.43	0.80
	MMW Filter	27.56	0.861
	LR-MMW Filter	31.65	0.91

Thirdly, the proposed LR-MMW Filter is evaluated to validate its efficacy for real-time images which are randomly collected based on a real-time environment. Table 3 mentioned below shows the efficiency of the proposed LR-MMW filter with the existing Wiener, Kalman and MMW filters by computing the value of PSNR and SSIM.

Table 3. Performance of various filters for the images obtained from real-time environment

Dataset	Filters	PSNR (dB)	SSIM
Images collected from real-time environment	Wiener Filter	25.93	0.745
	Kalman Filter	26.67	0.711
	MMW Filter	30.26	0.786
	LR-MMW Filter	33.71	0.831

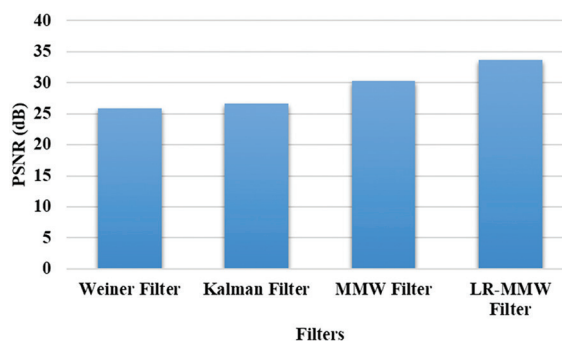


Fig. 9. Graphical representation of PSNR value for real-time images

The results from Table 3 and Fig. 9 show that the proposed LR-MMW filter is even effective for the data obtained from a real-time environment. While evaluating the PSNR value, the LR-MMW filter obtained a better PSNR value of 33.71 dB which is better than the existing Wiener, Kalman and MMW filters with 25.93 dB, 26.67dB, 30.26Db and 33.71dB respectively.

Overall, the LR-MMWF provides better PSNR and SSIM values DIV2K and BDS100 and the image obtained from a real-time environment, the better result is due to the utilized fidelity term and the regularization term in the image reconstruction model, and it increases the quality of the image and de-noising the region of background.

4.2. COMPARATIVE ANALYSIS

This section provides a comparison among various SR methodologies namely the Filter Net [16], 1D,2D-CNN filter [20] and the proposed LR-MMW filter. The values of PSNR and SSIM are evaluated for the BSD100 dataset based on the overall performance in providing SR images. The results obtained from the comparison show that the proposed LR-MMW attained better performance in the reconstruction of super-resolution images from the composed LR image. The LR-MMW method utilizes a multi-frame SR model which consists of fidelity terms (L_2 Norm) and regularization term (BTV prior). Moreover, the image obtained from the pro-

posed method attains better resolution than the existing methods. The graphical representation of PSNR and SSIM of LR-MMW with existing filter methods are represented in Fig. 9 and Fig. 10 respectively. Moreover, the comparative results for PSNR value and SSIM is shown in Table. 3 and Table. 4 respectively.

Table 4. Comparative table for PSNR

Filter methods	PSNR (dB)
1D,2D-CNN Filter [17]	31.63
Filter Net [21]	28.95
LR-MMW	31.65

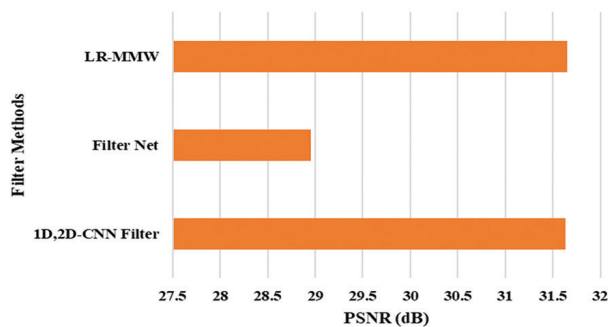


Fig. 10. Graphical representations of PSNR value

Table 4 and Fig. 10, it is shown that the proposed LR-MMW filter attains a high PSNR value of 31.65dB whereas the existing Filter Net and 1D, 2D CNN filter achieved PSNR values of 28.95 dB and 31.63 dB respectively. The presence of (L_2 Norm) and (BTV prior) present in the image reconstruction model help to obtain better PSNR values than the existing methodologies. This result is due to the presence of an image reconstruction model which consists of (L_2 Norm) and (BTV prior). The L_2 Norm reduces residues in the predicted HR image and BTV improves the stable state of generated HR image.

Table 5. Comparative table for SSIM

Filter methods	SSIM
1D,2D-CNN Filter [17]	0.726
Filter Net [21]	0.892
LR-MMW	0.914

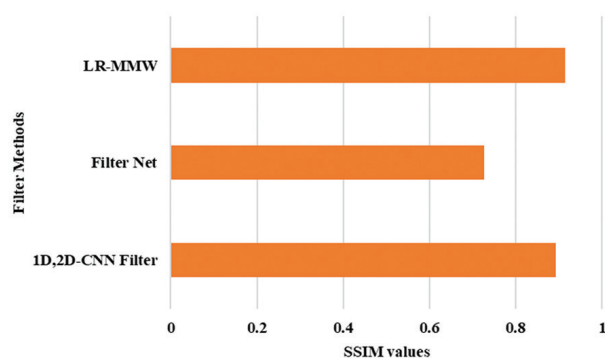


Fig. 11. Graphical representations of SSIM value

The above table.4 shows that the proposed LR-MMW attains higher structural similarity (SSIM) Of 0.914 whereas the existing 1D,2D CNN filter, and Filter Net achieved SSIM of 0.892 and 0.726 respectively. The graphical representation of SSIM values is represented in Fig. 11.

From Table.4 and Table.5, it is concluded that the performance of the LR-MMW filter provides better PSNR and SSIM values compared with the existing filter methods. The LR-MMW filters the residues in the predicted HR image using L_2 Norm, and the stable state of the HR image is determined using the BTV model. Thus LR-MMW filter provides better performance by improving the image quality and de-noising the background region effectively.

Moreover, the results are evaluated based on the spatial interpolation of the images obtained from Filter Net and LR-MMW filter. The images from the BSD 100 dataset with 3x upscaling are considered for evaluating the images whose sizes are 32x32, 64x64 and 128x128. Table 6 presents the outcome for varying image sizes for Filter Net and LR-MMW filter.

Table 6. Comparison of PSNR and SSIM by varying the pixel size of the images from the BSD100 dataset

Dataset	Size of the Image	PSNR (dB)		SSIM	
		Filter Net [21]	LR-MMW	Filter Net [21]	LR-MMW
BSD100 (3xupscaling)	32x32	28.45	33.78	0.7885	0.8128
	64x64	27.01	31.04	0.7179	0.7967
	128x128	27.05	30.97	0.7181	0.7342

The results from Table 6 show that the proposed LR-MMW achieved better interpolation for various image sizes obtained from BSD 100 dataset. For instance, the LR-MMW has obtained 31.04 Db for an image size of 64x64 which is comparatively higher than the existing Filter Net method. Similarly, SSIM value of LR-MMW is 0.7967 whereas Filter Net obtained 0.7179. These results prove the efficacy of the proposed LR-MMW filter and this better result is due to the bicubic interpolation which uses the 4x4 neighborhood method to determine the output.

5. CONCLUSION

This research proposed an LR-MMW filter to enhance the effectiveness of SR images to maximize both the analysis and human interpretation process. The results are evaluated based on the data collected from DIV2K and BSD 100 datasets, this research considered PSNR and SSIM to compute the efficiency of the proposed method. The benefit of employing the LR-MMW filter technique is that the edge signal is better retained when compared to the usage of the median and Wiener filter techniques. This helps to improve the quality of the deteriorated image using the drop-off effect.

Regarding the de-noising effect, the LR-MMW filter technique can significantly outperform traditional filters. The performance of the proposed LR-MMW filter is evaluated with the Wiener, Kalman and MMW filters for both DIV2K and BSD 100 datasets. For DIV2K, the proposed LR-MMWF obtained PSNR of 31.69 dB and SSIM of 0.92 and for BSD 100 the value of PSNR and SSIM is 31.65dB and 0.91 respectively which is comparatively higher than the existing filter techniques. In the future, the proposed model can be implemented with deep learning techniques to obtain better results.

6. REFERENCES

- [1] M. M. Khattab, A. M. Zeki, A. A. Alwan, B. Bouallegue, S. S. Matter, A. M. Ahmed, "Regularized Multiframe Super-Resolution Image Reconstruction Using Linear and Nonlinear Filters", *Journal of Electrical and Computer Engineering*, Vol. 2021, 2021, p. 8309910.
- [2] S. W. Remedios, S. Han, Y. Xue, A. Carass, T. D. Tran, D. L. Pham, J. L. Prince, "Deep filter bank regression for super-resolution of anisotropic MR brain images", *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Singapore, 18-22 September 2022, Vol. 13436, pp. 613-622.
- [3] G. Suryanarayana, K. Chandran, O. I. Khalaf, Y. Alo-taibi, A. Alsufyani, S. A. Alghamdi, "Accurate Magnetic Resonance Image Super-Resolution Using Deep Networks and Gaussian Filtering in the Stationary Wavelet Domain", *IEEE Access*, Vol. 9, 2021, pp. 71406-71417.
- [4] M. V. Daithankar, S. D. Ruikar, "Analysis of the Wavelet Domain Filtering Approach for Video Super-Resolution", *Engineering, Technology & Applied Science Research*, Vol. 11, No. 4, 2021, pp. 7477-7482.
- [5] W. Xu, H. Song, K. Zhang, Q. Liu, J. Liu, "Learning lightweight Multi-Scale Feedback Residual network for single image super-resolution", *Computer Vision and Image Understanding*, Vol. 197-198, 2020, p. 103005.
- [6] V. Curcio, L. A. Alemán-Castañeda, T. G. Brown, S. Brasselet, M. A. Alonso, "Birefringent Fourier filtering for single molecule coordinate and height super-resolution imaging with dithering and orientation", *Nature Communications*, Vol. 11, 2020, p. 5307.
- [7] K. Jonghyun, G. Li, I. Yun, C. Jung, J. Kim, "Edge and identity preserving network for face super-resolution", *Neurocomputing*, Vol. 446, 2021, pp. 11-22.
- [8] Y. Huang, J. Li, X. Gao, Y. Hu, W. Lu, "Interpretable Detail-Fidelity Attention Network for Single Image Super-Resolution", *IEEE Transactions on Image Processing*, Vol. 30, 2021, pp. 2325-2339.
- [9] I. Taghavi, S. B. Andersen, C. A. V. Hoyos, M. Schou, F. Gran, K. L. Hansen, M. B. Nielsen, C. M. Sørensen, M. B. Stuart, J. A. Jensen, "Ultrasound super-resolution imaging with a hierarchical Kalman tracker", *Ultrasonics*, Vol. 122, 2022, p. 106695.
- [10] S. Aymaz, C. Köse, "A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion", *Information Fusion*, Vol. 45, 2019, pp. 113-127.
- [11] Y. Chen, L. Liu, V. Phonevilay, K. Gu, R. Xia, J. Xie, Q. Zhang, K. Yang, "Image super-resolution reconstruction based on feature map attention mechanism", *Applied Intelligence*, Vol. 51, No. 7, 2021, pp. 4367-4380.
- [12] X. Feng, J. Li, Z. Hua, "Guided filter-based multi-scale super-resolution reconstruction", *CAAI Transactions on Intelligence Technology*, Vol. 5, No. 2, 2020, pp. 128-140.
- [13] F. Salvetti, V. Mazzia, A. Khaliq, M. Chiaberge, "Multi-image super resolution of remotely sensed images using residual attention deep neural networks", *Remote Sensing*, Vol. 12, No. 14, 2020, p. 2207.
- [14] Y. Shi, H. Zhong, Z. Yang, X. Yang, L. Lin, "DDet: Dual-Path Dynamic Enhancement Network for Real-World Image Super-Resolution", *IEEE Signal Processing Letters*, Vol. 27, 2020, pp. 481-485.
- [15] J. Cai, B. Huang, "Super-Resolution-Guided Progressive Pansharpening Based on a Deep Convolutional Neural Network", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 6, 2021, pp. 5206-5220.
- [16] J. Yang, Y.-Q. Zhao, J. C.-W. Chan, L. Xiao, "A Multi-Scale Wavelet 3D-CNN for Hyperspectral Image Super-Resolution", *Remote Sensing*, Vol. 11, No. 13, 2019, p. 1557.
- [17] J. Park, J. Lee, D. Sim, "Low-complexity CNN with 1D and 2D filters for super-resolution", *Journal of*

- Real-Time Image Processing, Vol. 17, No. 6, 2020, pp. 2065-2076.
- [18] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, Y. Fu, "LatticeNet: Towards Lightweight Image Super-Resolution with Lattice Block", *Computer Vision – ECCV 2020, Lecture Notes in Computer Science*, Vol. 12367, Springer International Publishing, 2020, pp. 272-289.
- [19] A. Esmailzadeh, M. O. Ahmad, M. N. S. Swamy, "MuRNet: A deep recursive network for super resolution of bicubically interpolated images", *Signal Processing: Image Communication*, Vol. 94, 2021, p. 116228.
- [21] Y. Chen, L. Liu, V. Phonevilay, K. Gu, R. Xia, J. Xie, Q. Zhang, K. Yang, "Image super-resolution reconstruction based on feature map attention mechanism", *Applied Intelligence*, Vol. 51, No. 7, 2021, pp. 4367-4380.
- [22] F. Li, H. Bai, Y. Zhao, "FilterNet: Adaptive Information Filtering Network for Accurate and Fast Image Super-Resolution", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 30, No. 6, 2020, pp. 1511-1523.
- [23] D. Gao, D. Zhou, "A very lightweight and efficient image super-resolution network", *Expert Systems with Applications*, Vol. 213, 2023, p. 118898.
- [24] Z. He, D. Chen, Y. Cao, J. Yang, Y. Cao, X. Li, S. Tang, Y. Zhuang, Z. Lu, "Single image super-resolution based on progressive fusion of orientation-aware features", *Pattern Recognition*, Vol. 133, 2023, p. 109038.
- [25] BSD100 dataset, <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html/dataset/images.html> (accessed: 2023)
- [26] DIV2K dataset, <https://data.vision.ee.ethz.ch/cvl/DIV2K/> (accessed: 2023)

JPEG2000-Based Semantic Image Compression using CNN

Original Scientific Paper

Anish Nagarsenker

Dr Vishwanath Karad MIT World Peace University,
School of Electronics and Communication Engineering
Kothrud, Pune, India
anishnagarsenker@gmail.com

Prasad Khandekar

Dr Vishwanath Karad MIT World Peace University,
School of Electronics and Communication Engineering
Kothrud, Pune, India
prasad.khandekar@mitwpu.edu.in

Minal Deshmukh

Vishwakarma Institute of Information and Technology,
Department of Electronics and Telecommunication Engineering
Pune, India
minal.deshmukh@viit.ac.in

Abstract – Some of the computer vision applications such as understanding, recognition as well as image processing are some areas where AI techniques like convolutional neural network (CNN) have attained great success. AI techniques are not very frequently used in applications like image compression which are a part of low-level vision applications. Intensifying the visual quality of the lossy video/image compression has been a huge obstacle for a very long time. Image processing tasks and image recognition can be addressed with the application of deep learning CNNs as a result of the availability of large training datasets and the recent advances in computing power. This paper consists of a CNN-based novel compression framework comprising of Compact CNN (ComCNN) and Reconstruction CNN (RecCNN) where they are trained concurrently and ideally consolidated into a compression framework, along with MS-ROI (Multi Structure-Region of Interest) mapping which highlights the semiotically notable portions of the image. The framework attains a mean PSNR value of 32.9dB, achieving a gain of 3.52dB and attains mean SSIM value of 0.9262, achieving a gain of 0.0723dB over the other methods when compared using the 6 main test images. Experimental results in the proposed study validate that the architecture substantially surpasses image compression frameworks, that utilized deblocking or denoising post-processing techniques, classified utilizing Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measures (SSIM) with a mean PSNR, SSIM and Compression Ratio of 38.45, 0.9602 and 1.75x respectively for the 50 test images, thus obtaining state-of-art performance for Quality Factor (QF)=5.

Keywords: Computer Vision, Neural Networks, CNN, MS-ROI, Compression, JPEG2000

1. INTRODUCTION

Image Compression is a form of data compression procedure that is used on digital images to reduce their size as well as cost and make them more suitable for storage or communication. Properties of Image Compression include Scalability (Component Progressive, Quality Progressive, and Resolution Progressive), Meta Information, Interest Coding, and Processing Power [1]. Lossy image compression standards attain a greater compression ratio in contrast to lossless im-

age compression standards, but there is a loss of image quality. Lossless image compression is mainly used in medical imaging, clip arts, technical drawings, or comics whereas Lossy image compression maybe used for natural images such as photographs [2, 3]. A subset of Machine Learning, Deep Learning algorithms aim to extract identical inferences as humans would, by continuously analyzing data with the specified analytical structure, and hence, the multiple-layered structures of algorithms are employed called neural networks [4].

In Deep learning, a class of ANN that is usually applied to image processing and image recognition to analyze visual imagery is called as Convolutional Neural Network (CNN). Some advantages of using neural networks include (a) producing output irrespective of the input provided to them because they learn by themselves (b) performing multiple tasks concurrently without affecting system performance (c) instead of a database, the input is stored in its own network and hence loss of data does not affect its working (d) it works through real-time events because it learns from events and acts accordingly when a similar event arises (e) ability of the network to detect the fault by themselves and produce the output accordingly [5, 6]. Numerous layers exist in CNN like fully connected layers, integration layers, intermittent layers, and convolutional layers whereby intermittent layers and integration layers have no variables associated with them whereas fully-connected layers and convolutional layers have variables associated with them [7].

ComCNN gains an excellent compact characterization from the image at input hence conserves the structural details of the image [8]. RecCNN has been utilized to rebuild the image with excellent quality, thus producing an output image of high quality and maintaining all the structural details of the input image [9]. The process of portioning every pixel in an image in its region that has semantic value with a specific label is called as semantic segmentation. MS-ROI mapping is a form of region-based semantic segmentation that can be used to successfully prime on localization functions irrespective of the amount of classes. MS-ROI mapping is used to represent numerous prominent areas of a picture in one pass and hence can be used to boost the visual attributes of only the semantic objects, thus making the image visually attractive [10].

PSNR and SSIM are two parameters that are widely used for the assessment of picture quality. The ratio of the highest plausible power from the signal provided to the highest viable power of the distorted noise which impacts the accuracy of characterization is called PSNR. In lossy compression, PSNR is widely used to check the restoration standard of videos and images, where PSNR is inversely proportional to Mean Squared Error (MSE). A perceptual metric that measures image caliber variation generated from rectification like data transmission or data compression losses is known as Structural Similarity Index (SSIM). SSIM values range between 0 and 1 whereas PSNR values range between 0 and ∞ [11].

In [9], a framework is proposed comprising of ComCNN and RecCNN, that are ideally consolidated into a compression framework to achieve image compression of superlative quality at comparatively lower bit rates using JPEG, giving high values of PSNR and SSIM at QF=5 and QF=10. In [10], using MSROI heat map, semantically salient regions of the image are highlighted using CNN and these are then encrypted at a higher caliber than the background portions of the image using JPEG

giving significantly high values of PSNR and SSIM. The framework utilised in this paper, is a combination of the methods used in these 2 papers giving a novel framework standard for Image Compression.

In this paper, a framework is proposed comprising 2 CNNs that are ideally consolidated into a compression framework along with MS-ROI (Multi Structure-Region of Interest) to achieve image compression of the highest quality at comparatively lower bit rates using JPEG2000 giving high values of PSNR and SSIM. The structural information is maintained by the Compact CNN (1st CNN) which grasps excellent compact characterization from the given image followed by which an MS-ROI map is generated where semantically salient regions are highlighted and passed through the JPEG2000 encoder. The decrypted image is reconstructed with high standards using Reconstruction CNN (2nd CNN) to obtain the final reconstructed image after compression. Concurrently learning ComCNN and RecCNN eases the precise reformation of the decrypted image using RecCNN [9]. When tested using 50 test images, this framework is able to achieve a mean PSNR, SSIM and Compression Ratio of 38.45, 0.9602 and 1.75x respectively, thus outperforming other image enhancement methods for QF=5. The following sections give a brief overview of the methods employed, methodology performed, results obtained, and conclusions procured from the results.

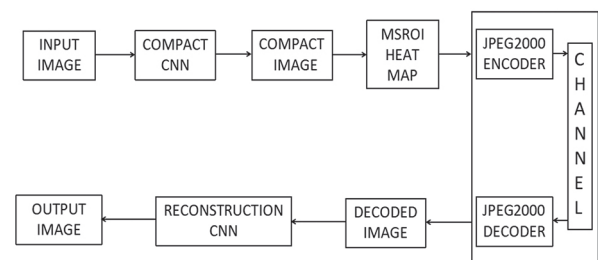


Fig. 1. Proposed Block Diagram

2. COMPACT CNN(COM-CNN) AND RECONSTRUCTION CNN(REC-CNN)

CNNs, which are composite feed-forward neural structures, because of their immense accuracy, are used in image recognition and classification, Segmentation and Object Detection [12]. The CNN follows a progressive model which deals with building an organization and finally provides a fully-associated layer where the neurons are associated with one another and the output is extracted. CNNs consist of the Convolutional layer in which filters are utilized for the input image or attribute maps. The integration layer is utilized to lower the scale of the attribute map and the Fully-Connected layer which are positioned prior to classification yield and is utilized to normalize the outcomes before classification [11].

ComCNN gains an excellent compact characterization from the image at input hence conserves the structural details of the image. ComCNN normally con-

sists of 3 layers compared to RecCNN which contains a higher number of layers (Around 20 layers) out of which, most layers undergo the process of convolution plus batch normalization plus ReLU. As such, RecCNN has been utilized to rebuild the image with excellent quality, thus producing an output image of high quality and maintaining all the structural details of the input image. ComCNN and RecCNN attempt to build the reconstructed image identical to the input image. Training both the CNNs concurrently yields better results as compared to training the CNNs individually [9].

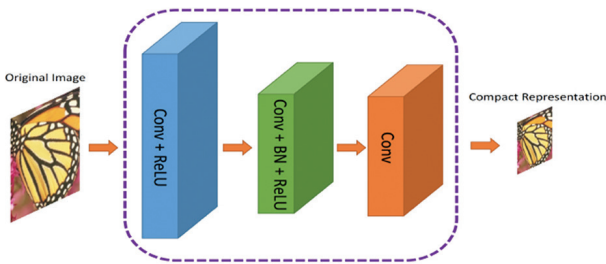


Fig. 2. ComCNN Architecture

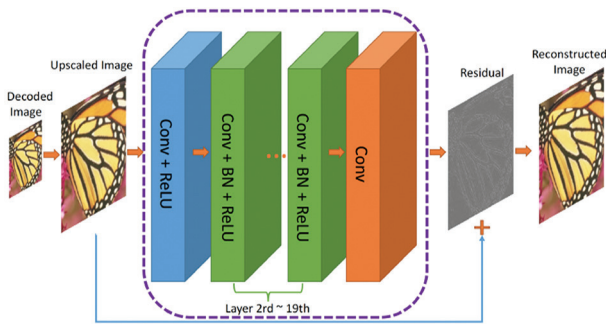


Fig. 3. RecCNN Architecture

3. MULTI-STRUCTURE REGION OF INTEREST (MS-ROI)

The process of portioning every pixel in an image in its region that has a semantic value with a specific label is called semantic segmentation. MS-ROI heat map is a method that stabilizes the activation for numerous entities and hence doesn't agonize from the concern of global average pooling. Unconstrained by the amount of classes, this method permits to productively train on localization tasks. MS-ROI is able to spot and roughly detect all the objects in the given compact image. In this framework, a group of 3D attribute plots is preferred in which each attribute plot is grasped for a discrete domain and is grasped autonomously for the plots for auxiliary domains. In MS-ROI, the number of classes is reduced by combining homogenous types of classes into a broader class as the aim is to associate all the objects present in the image. The plot created will be practically indistinguishable till the time the entities of these merged domains have homogenous shapes and are inside the identical universal type. By utilizing the MS-ROI heat map technique, similarity merit for each element in the domain $[0,1]$ is procured in which 1 specifies maxi-

mal similarity [3, 10]. This technique is able to highlight the semantically notable areas of the image which are then encrypted at a superior grade in contrast to the other less notable areas. Therefore, MS-ROI mapping is used to represent numerous prominent areas of a picture in one pass and hence can be used to increase the visual aspect of only the semantic objects in the picture, thus making the image visually attractive [10].

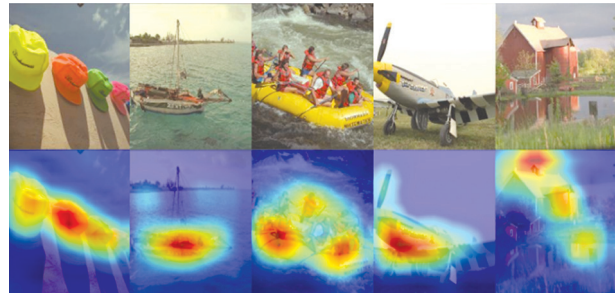


Fig. 4. MSROI Heat Map Examples

4. JPEG2000

JPEG 2000, invented by the Joint Photographic Experts Group, is an international-grade compression technique established on the concept of wavelet transform and provides an exceptionally superior level of robustness and accessibility. The prime advantages of utilizing this approach comprise intensified transmission and locale of interest coding. It is a strong and suitable mechanism that is specifically used in RS picture processing and capacity. It is an image coding system that provides lossy image compression for storage but can maintain an image quality similar to that of the original image [13].

In the JPEG-2000 method, the input image is given to a group of discrete wavelet filters which converts the pixel details into wavelet coefficients, which are then assembled into various bands. Every band in the image carries the wavelet coefficients that portray a discrete level and sharp structural frequency domain of the complete input image. These bands are then given for quantization where the bit streams, accommodated in each code block, are pruned (re-quantized). These bit streams are impeccably pruned utilizing a method called as PCRC (post-compression rate control). The code blocks are assembled conforming to their significance where the information related to the magnitude of bits is determined by a process called context modeling [13, 14].

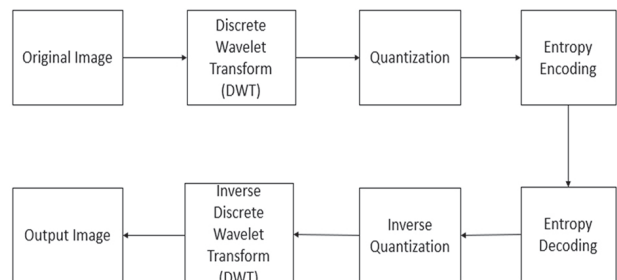


Fig. 5. JPEG2000 Encoder and Decoder

In the decoder section, the bit streams are given for entropy decoding followed by inverse quantization, where the bits are decoded according to their significance. These are then driven through a group of inverse discrete wavelet filters where the wavelet coefficients are transformed back to pixels. The decoded image is obtained which contains pixels according to their significance where the most important contents are observed earlier than the other contents. The final image obtained is a compressed variant of the actual image with very little loss in image quality [15].

5. METHODOLOGY

Image Acquisition

In this phase, an input image is given to the proposed model. The pre-processed and normalized image should be either RGB or Grayscale. The complete dataset is split into 2 sets (training and testing) with the 70-30% data splitting or 10-fold cross-validation method.

Compact CNN

This compact solution minimises the number and sophistication of lower-level kernels in a CNN by separating the colored information of the input picture. It is presented a compact construction of fully convolutional neural networks including worldwide color histogram features. The design proposed decreases the number of processors required to extract high-level given grayscale input. The compact structure has 40% fewer parameters to tune and performs comparably to the standard CNN formed on ROB pictures. In the first layer, two CNNs with variable amounts of kernels are trained: an original title and a compact counterpart. The 'original' network is a carbon copy of CNN. When providing the final results for both the regular Network and the compressed CNN, however, we only use single-view testing. Both the traditional and compact CNNs include four convolution layers. The condensed CNN's first stratum has just 32 kernels, while the traditional CNN's first layer contains 64 kernels. This cuts the number of variables in layer 3 (the intermediate convolution layer) by half, resulting in a compacted CNN with 40% lower dimensionality to tweak than the previous form. The concatenation is denoted as:

$$y^{j(r)} = \text{ReLU}(b^{j(r)} + \sum_i k^{ij(r)} * x^{i(r)}) \quad (1)$$

MSROI

The objective of ROI (Region of Interest) involves isolating a section of an image that we want to refine or manipulate. To start with, a binary mask is generated which is basically a binary image, the same size as the image we want to examine, with all other pixels set to 0 and only the ones forming the ROI set to 1. Pre-processing (Edge detection), Polygonal approximation, and Shape signature extraction are all factors to consider while extracting MS-ROI from a compact picture. The steps pursued in the implementation of MS-ROI

include defining a group of boundary pixels analogous to the ROI, converting every pixel in the set to a parametric curve in the parameter space, supplementing the pixels in the accumulator matrix 'A' as guided by the parameter curve, evaluating the provisional feature vector, vector mean, and feature vector normalization.

JPEG Encoder and Decoder

The JPEG encoder and decoder provide an inbuilt library to achieve the functionality of compression. The steps followed in JPEG encoding and decoding include taking the picture to encode, getting the resulting image encoded and decoded by the program, choosing the quantization coefficient and quantization matrix for luminance, viewing the details of the block (square of 64 pixels) and size of the image and then, passing on the resulting image to the RecCNN.

Reconstruction CNN

Using a revised DAS algorithm (back projection-based) $D: R_m R_n$, the technique used first computes a low-graded appraisal of x , indicated by \bar{x} using measurement recorded with a single graphical representation. We define DWH, where $W: R_n R_n$ is the i th transmitter's "pixel-wise" re-evaluation operator (diagonal matrix).

$$w_{i(r)} = \left(\overline{h_i^x}(r) \sum_{j=1}^{n_r} \overline{h_j^x}(r) \right)^{-1} \quad (2)$$

As post-DAS picture equalisation, it compensates for the amplitude-related impacts of far-field diffraction. A CNN $f: R_n R_n$, trained to retrieve a high-quality estimate of x as $x=f(x)$, is given the approximation $x=Dy$ in the second stage. The use of a CNN for restoration jobs is recommended over employing end-to-end techniques that attempt to directly map a measurement space to an image space. We want to train a mapping f from a low-quality picture subspace $W R_n$ to a high-quality image subspace $V R_n$. To define these subspaces more accurately, a transducer with a specific aperture is used, made of an array of transducer components with a given shape, center frequency, and bandwidth. To reassemble US pictures using D , we define W as the subspace of measurements recorded by a single in sonification utilising the whole aperture. High SL and EW artefacts, as well as probable GL artefacts, may be seen in these photos (e.g., linear-array designs). In order to avoid GLs, a transducer with the same aperture and physical parameters as the one used for W is suggested for V . To create reference pictures from this array, we use the D operator for each in sonification and coherent compounding. SA is commonly regarded as the gold standard, generating pictures with great resolution while reducing SL and EW artefacts.

Moreover, SA photographs have better speckle patterns than "exotic" ground truth images. We ensure that both subspaces of the CNN include speckle patterns from sub-resolution scattering interferences. Due to the assumptions made in the physical measurement technique (2) and the following back projection opera-

tor (3) used to determine W and V , the trained CNN is not expected to repair artefacts from ignored physical events. With increased resolution, diffraction artefacts are reduced yet speckle is retained.

6. RESULTS AND DISCUSSION

The execution of the proposed algorithm is assessed using comparisons between the SSIM and PSNR of various post-processing denoising methods as well as de-blocking methods for images such as Zhang's, DicTV, Ren's, Sun's, WNNM, ARCNN and BM3D because of their excellent performance in producing output images. Also, the ComCNN and RecCNN are individually trained as well as trained simultaneously for comparison with the PSNR and SSIM values obtained using the proposed framework. The values of PSNR and SSIM obtained from the combination of ComCNN and RecCNN together, outperforms the values of PSNR and SSIM obtained using the 2 CNNs individually as seen from Tables I and II [9]. This result showcases the importance of training the 2 CNNs simultaneously and using them along with MS-ROI heat map, which represents numerous prominent areas of a picture in one pass and hence can be used to increase visual aspect of only the semantic objects in the picture, making the image visually attractive and thus giving excellent values of PSNR as well as SSIM for QF=5. In the experiments performed, the 2 CNNs (ComCNN and RecCNN) are trained simultaneously using a set of approximately 1000 images (256 x 256) of various different types like vehicles, people, scenery, plants, animals etc. and is trained for 200 epochs. The experiments are executed in Python Environment (Anaconda) on Laptop having Intel(R) Core (TM) i5-10210U CPU @1.60GHz 2.11 GHz and an Nvidia GeForce MX230 GPU. Around 12-15 hours of PC time is required for training the CNNs simultaneously up to 200 epochs on GPU [9].

In the experiments performed, 6 test images as that used in the paper [9] are used, so that the comparison becomes easier with all the Image Enhancement methods. From the PSNR and SSIM table for QF=5, it can be concluded that the framework attains PSNR values in

the domain 31.28-35.26dB with a mean PSNR value of 32.9dB, achieving a gain of 3.52dB and achieves SSIM values in the domain 0.9065-0.9538dB with a mean SSIM value of 0.9262, achieving a gain of 0.0723dB over the "ComCNN+RecCNN [9]" method which gives optimal outputs of PSNR and SSIM in comparison to every method mentioned in Table 1 and 2. The framework proposed eclipses each and every image intensification method including the ARCNN [20], which is considered the landmark of CNN. The framework proposed not only eliminates a bulk of the antiquities notably but also maintains more attributes on the texture as well as the edges. This framework is able to achieve additional high-frequency details, retrieve salient edges as well as perfect textures in the reproduced image.

A total of 50 test images were used and some of the output images along with their outputs are displayed in Fig. 9 and Table 3 respectively. Various test images with QF=5 are shown in Table III and their corresponding input size (in KB), output size (in KB), Compression Ratio, PSNR, and SSIM are noted down. For these test images, the Compression Ratio values range between 1.23x-3.15x with an average compression rate of 1.62x, the PSNR values range between 31.28-45.37dB with an average PSNR of 37.26dB while the SSIM values range between 0.8889-0.9915dB with an average SSIM of 0.9556dB. These values of PSNR and SSIM suggest that the framework proposed for compression undergoes very minute changes from the input image to the output image with a considerable compression rate and hence, an output image is obtained which is very similar to the input image.

As can be seen from the pictures displayed in Fig. 6 and Fig. 9, there is not much visual difference observed between the input and the output images (the output image is the same as the input image) with a significant amount of compression achieved. This is because the MS-ROI heat map highlights the semantically notable areas of the image which are then encrypted at a superior grade in contrast to the other less notable areas and the 2 CNNs work simultaneously to rebuild

Table 1. Analogy of PSNR of the test inputs of various Image deblocking, denoising and compression methods for QF=5 with the framework proposed

TEST IMAGES	Parrots	Cameraman	Lena	Butterfly	House	Peppers	Mean
	PSNR						
JPEG	26.19	24.45	27.33	22.58	27.77	27.17	25.92
BM3D [11]	27.33	25.27	28.63	24.05	29.21	28.52	27.17
Ren's [10]	27.87	25.46	29.07	24.58	29.66	29.07	27.62
Zhang's [12]	27.78	25.39	29.00	24.20	29.24	29.07	27.45
Sun's [4]	27.45	25.25	28.87	23.83	29.09	29.05	27.26
DicTV [9]	26.83	24.54	28.07	23.10	28.45	27.95	26.49
Zhang's [13]	28.27	25.61	29.51	25.30	30.12	29.61	28.07
ARCNN [14]	28.13	25.27	29.31	25.64	29.68	29.02	27.84
WNNM [15]	27.80	25.49	28.95	24.75	29.62	28.99	27.60
ComCNN [5]	27.67	24.93	29.46	23.05	29.17	29.33	27.27
RecCNN [5]	28.52	26.33	29.63	25.99	30.13	29.81	28.40
ComCNN+	30.12	26.53	31.14	26.23	31.45	30.84	29.38
RecCNN [5]							
Proposed	32.58	33.19	33.28	31.28	31.79	35.26	32.90

Table 2. Analogy of SSIM of the test inputs of various Image deblocking, denoising and compression methods for QF=5 with the framework proposed

TEST IMAGES	Parrots	Cameraman	Lena	Butterfly	House	Peppers	Mean
	SSIM						
JPEG	0.7581	0.7283	0.7367	0.7378	0.7733	0.7087	0.7404
BM3D [11]	0.8118	0.7607	0.7837	0.8184	0.8082	0.7639	0.7911
Ren's [10]	0.8310	0.7666	0.8010	0.8419	0.8197	0.7876	0.8080
Zhang's [12]	0.8308	0.7672	0.8035	0.8313	0.8141	0.7895	0.8061
Sun's [4]	0.8323	0.7687	0.8061	0.8321	0.8113	0.7931	0.8073
DicTV [9]	0.8005	0.6658	0.7744	0.7769	0.7963	0.7456	0.7599
Zhang's [13]	0.8460	0.7666	0.8169	0.8667	0.8285	0.8031	0.8213
ARCNN [14]	0.8446	0.7674	0.8142	0.8741	0.8209	0.7961	0.8196
WNNM [15]	0.8287	0.7674	0.7947	0.8445	0.8178	0.7827	0.8060
ComCNN [5]	0.8377	0.7662	0.8042	0.7488	0.8119	0.7966	0.7942
RecCNN [5]	0.8497	0.7945	0.8195	0.8760	0.8251	0.8004	0.8275
ComCNN+	0.8951	0.8167	0.8486	0.8847	0.8456	0.8328	0.8539
RecCNN [5]							
Proposed	0.9307	0.9065	0.9538	0.9155	0.9189	0.9315	0.9262

Table 3. PSNR, SSIM and Compression Ratios of the various test images for QF=5

Test Inputs	Input Size(KB)	Output Size(KB)	Compression Ratio	PSNR	SSIM
Peppers	6.01	4.12	1.46x	35.26	0.9315
Parrots	6.20	4.32	1.44x	32.58	0.9307
House	4.56	3.69	1.24x	31.79	0.9189
Lena	7.81	5.39	1.45x	33.28	0.9538
Butterfly	10.2	5.40	1.89x	31.28	0.9155
Cameraman	6.81	4.66	1.46x	33.19	0.9065
TT1	760	499	1.52x	37.7	0.9813
TT2	771	472	1.63x	38.47	0.9905
TT3	333	138	2.41x	45.1	0.9828
TT4	194	155	1.25x	43.27	0.9915
TT5	711	550	1.29x	45.37	0.9875
TT6	1400	880	1.59x	35.96	0.9812
TT7	752	606	1.24x	38.81	0.9887
TT8	4530	1440	3.15x	37.98	0.8889
TT9	1540	1250	1.23x	38.82	0.9845
Mean	-	-	1.62x	37.26	0.9556

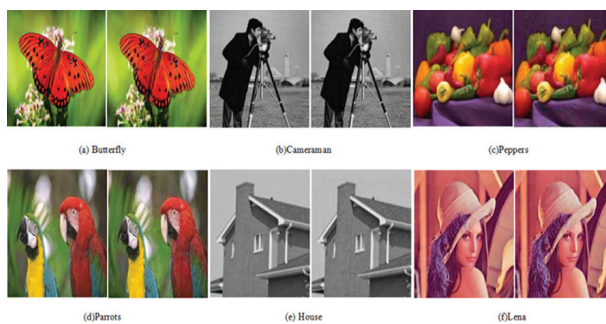


Fig. 6. Different Test Images and their Outputs

the original image, generating an image appearing similar to the original image, hence giving higher values of PSNR as well as SSIM. The greater the number of epochs used in the original image, hence giving higher values of PSNR as well as SSIM. The greater the number of epochs used in the training of CNNs concurrently, the better results the framework produces. Overall, a total of 50 test images were used to validate the quality of framework and the results show that this framework achieves a mean PSNR, SSIM and Compression Ratio of 38.45, 0.9602 and 1.75x respectively, thus proving to be an ideal framework for image compression.

The graphs in Fig. 7 show the mean PSNR and SSIM obtained for the 6 test inputs using various Image compression, deblocking, and denoising methods. From the graphs, it can be concluded that the mean PSNR and SSIM obtained for the 6 test inputs at QF=5 easily outperforms all the other image compression, deblocking, and denoising methods to achieve significantly high values of 32.90 and 0.9262 respectively.

The graphs in Fig. 8 show the PSNR, SSIM, and Compression Ratio values obtained for the various test images, used in this experiment of image compression using CNNs. From the graphs, it can be concluded that the mean PSNR of 37.26, mean SSIM of 0.9556, and an average compression of 1.62x for the test images at QF=5 make this framework ideal for use. The PSNR values obtained for the test images in Fig. 9 are greater than 30 while the SSIM values are greater than 0.88 which shows a significant improvement over other image compression as well as image deblocking and denoising methods. These graphs in Fig. 7 and 8 are direct evidence that the proposed framework gives an output image that undergoes very minute changes compared to the input image along with significant compression, achieved in the range of 1.18x-3.75x.

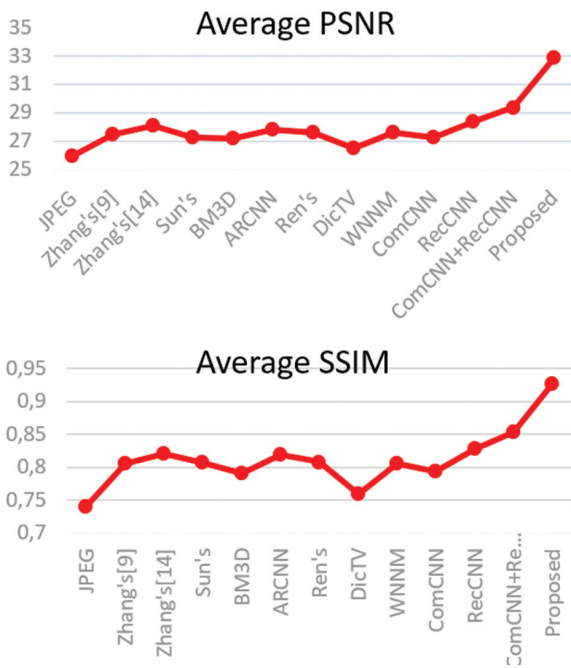


Fig. 7. Analogy of the mean PSNR and SSIM with various image compression, deblocking and denoising methods for QF=5

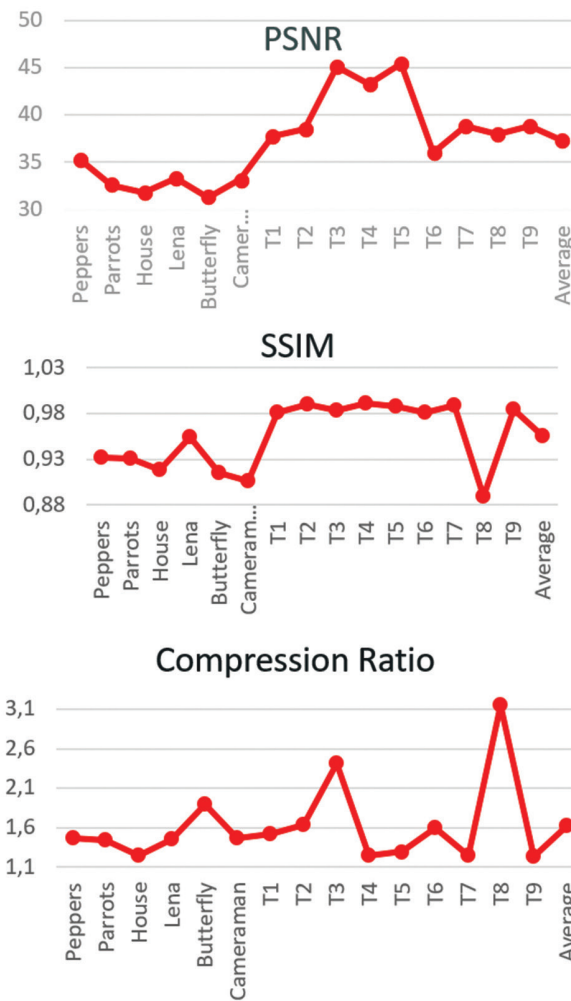


Fig. 8. Analogy of the mean PSNR, SSIM and Compression Ratio for the 15 Test Images for QF=5

7. CONCLUSION AND FUTURE WORK

The project consists of a novel compression framework involving CNN comprising of ComCNN and RecCNN where they are trained concurrently and ideally consolidated into a compression framework, along with MS-ROI mapping which highlights the semiotically notable portions of the image and JPEG2000 image codec for image compression. ComCNN gains an excellent compact characterization from the image at the input, hence conserving the structural details of the image, RecCNN has been utilized to rebuild the image with excellent quality, thus producing an output image of high quality and maintaining all the structural details of the input image whereas MS-ROI mapping is used to represent numerous prominent areas of a picture in one pass and hence can be used to boost the visual attributes of only the semantic objects, thus making the image visually attractive. The results of the experiments performed depict that this framework attains excellent values of PSNR and SSIM for QF=5, outperforming all the post-processing innovations, providing a suitable compression ratio of 1.18-3.75 times the original image and a mean PSNR and SSIM value obtained as 38.45 and 0.9602 respectively. For future work, the same framework can be used for examining the outputs (PSNR and SSIM) with different values of QF. Secondly, different architectures of CNNs can be tested for providing improved performances over the proposed method. Also, different image compression methods can be applied to this framework like JPEG, BPG, etc to check for their performance in comparison to the proposed JPEG2000 image.

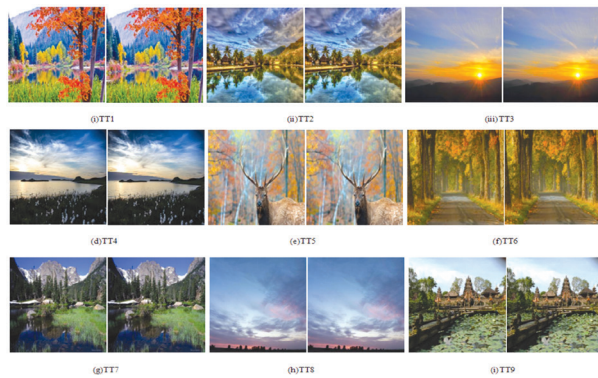


Fig. 9. Different Test Inputs And Their Outputs

8. REFERENCES

- [1] D. Mishra, S. K. Singh, R. K. Singh, "Deep Architectures for Image Compression: A Critical Review", Signal Processing, Vol. 191, 2022.
- [2] M. Gnacek, C. Heatwole, D. Fan, M. Hoffman, "Convolutional Neural Networks for Enhanced Compression Techniques", Proceedings of the IEEE National Aerospace and Electronics Conference, Dayton, OH, USA, 16-19 August 2021.

- [3] Cheng Wang, Yifei Han, Weidong Wang, "An End-to-End Deep Learning Image Compression Framework Based on Semantic Analysis", *Applied Sciences*, Vol. 9, No. 17, 2019.
- [4] Z. Zhang, G. Xu, J. Song, "CubeSat cloud detection based on JPEG2000 compression and deep learning", *Advances in Mechanical Engineering*, Vol. 10, No. 10, 2018.
- [5] Z. Liu, T. Liu, W. Wen, L. Jiang, J. Xu, Y. Wang, G. Quan, "DeepN-JPEG: A Deep Neural Network Favorable JPEG-based Image Compression Framework", *Proceedings of the 55th Design Automation Conference*, San Francisco, CA, USA, 24-29 June 2018, pp. 1-6.
- [6] O. H. Nagoor, J. Whittle, J. Deng, B. Mora, M. W. Jones, "Lossless Compression for Volumetric Medical Images using Deep Neural Network with Local Sampling", *Proceedings of the IEEE International Conference on Image Processing*, Abu Dhabi, UAE, 25-28 October 2020.
- [7] D. Sun, W.-K. Cham, "Postprocessing of low bit-rate block DCT coded images based on a field of experts prior", *IEEE Transactions on Image Processing*, Vol. 16, No. 11, 2007, pp. 2743-2751.
- [8] H. Fu, F. Liang, B. Lei, N. Bian, Q. Zhang, M. Akbari, J. Liang, C. Tu, "Improved Hybrid Layered Image Compression using Deep Learning and Traditional Codecs", *Signal Processing: Image Communication*, Vol. 82, 2020.
- [9] F. Jiang, W. Tao, S. Liu, J. Ren, X. Guo, D. Zhao, "An End-to-End Compression Framework Based on Convolutional Neural Networks", *Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology*, Snowbird, UT, USA, 2 August 2017.
- [10] A. Prakash, N. Moran, S. Garber, A. DiLillo, J. Storer, "Semantic Perceptual Image Compression using Deep Convolution Networks", *Proceedings of the Data Compression Conference*, 4-7 April 2017, pp. 250-259.
- [11] H. M. Yasin, A. M. Abdulazeez, "Image Compression Based on Deep Learning: A Review", *Asian Journal of Research in Computer Science*, Vol. 8, No. 1, 2021, pp. 62-76.
- [12] A. Ramdani, A. Virgono, C. Setianingsih, "Food Detection with Image Processing using Convolutional Neural Network (CNN) method", *Proceedings of the IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology*, Bali, Indonesia, 7-8 July 2020.
- [13] E. Jalilian, H. Hofbauer, A. Uhl, "Iris Image Compression Using Deep Convolutional Neural Networks", *Sensors*, Vol. 22, No. 7, 2022.
- [14] J. Sarkhawas, P. D. Khandekar, A. A. Kulkarni, "Variable Quality Factor JPEG Image Compression Using Dynamic Partial Reconfiguration and Microblaze", *Proceedings of the International Conference on Computing Communication Control and Automation*, Pune, India, 26-27 February 2015, pp. 620-624.
- [15] H. Chang, M. K. Ng, T. Zeng, "Reducing artifacts in jpeg decompression via a learned dictionary", *IEEE Transactions on Signal Processing*, Vol. 62, No. 3, 2014, pp. 718-728.
- [16] J. Ren, J. Liu, M. Li, W. Bai, Z. Guo, "Image blocking artifacts reduction via patch clustering and low-rank minimization", *Proceedings of the Data Compression Conference*, Snowbird, UT, USA, 20-22 March 2013, p. 516.
- [17] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering", *IEEE Transactions on Image Processing*, Vol. 16, No. 8, 2007, pp. 2080-2095.
- [18] M. Gnacek, C. Heatwole, D. Fan, M. Hoffman, "Convolutional Neural Networks for Enhanced Compression Techniques", *Proceedings of the IEEE National Aerospace and Electronics Conference*, Dayton, OH, USA, 16-19 August 2021.
- [19] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, W. Gao, "Concolor: constrained non-convex low-rank model for image deblocking", *IEEE Transactions on Image Processing*, Vol. 25, No. 3, 2016, pp. 1246-1259.
- [20] C. Dong, Y. Deng, C. C. Loy, X. Tang, "Compression artifacts reduction by a deep convolutional network", *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 7-13 December 2015, pp. 576-584.
- [21] S. Gu, L. Zhang, W. Zuo, X. Feng, "Weighted nuclear norm minimization with application to image denoising", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23-28 June 2014, pp. 2862-2869.

Enhancing Dynamic Hand Gesture Recognition using Feature Concatenation via Multi-Input Hybrid Model

Original Scientific paper

Djazila Souhila Korti

Belhadj Bouchaib University of Ain-Temouchent
Smart Structures Laboratory (SSL)
Faculty of Technology, Department of Telecommunication
Ain-Temouchent, Algeria
souhila.korti@univ-temouchent.edu.dz

Zohra Slimane

Abou Bekr Belkaid University of Tlemcen
Faculty of Technology, Department of Telecommunication
Tlemcen, Algeria
zoh_slimani@yahoo.fr

Kheira Lakhdari

Abou Bekr Belkaid University of Tlemcen
Faculty of Technology, Department of Telecommunication
Tlemcen, Algeria
kblakhdari@gmail.com

Abstract – Radar-based hand gesture recognition is an important research area that provides suitable support for various applications, such as human-computer interaction and healthcare monitoring. Several deep learning algorithms for gesture recognition using Impulse Radio Ultra-Wide Band (IR-UWB) have been proposed. Most of them focus on achieving high performance, which requires a huge amount of data. The procedure of acquiring and annotating data remains a complex, costly, and time-consuming task. Moreover, processing a large volume of data usually requires a complex model with very large training parameters, high computation, and memory consumption. To overcome these shortcomings, we propose a simple data processing approach along with a lightweight multi-input hybrid model structure to enhance performance. We aim to improve the existing state-of-the-art results obtained using an available IR-UWB gesture dataset consisting of range-time images of dynamic hand gestures. First, these images are extended using the Sobel filter, which generates low-level feature representations for each sample. These represent the gradient images in the x-direction, the y-direction, and both the x- and y-directions. Next, we apply these representations as inputs to a three-input Convolutional Neural Network- Long Short-Term Memory- Support Vector Machine (CNN-LSTM-SVM) model. Each one is provided to a separate CNN branch and then concatenated for further processing by the LSTM. This combination allows for the automatic extraction of richer spatiotemporal features of the target with no manual engineering approach or prior domain knowledge. To select the optimal classifier for our model and achieve a high recognition rate, the SVM hyperparameters are tuned using the Optuna framework. Our proposed multi-input hybrid model achieved high performance on several parameters, including 98.27% accuracy, 98.30% precision, 98.29% recall, and 98.27% F1-score while ensuring low complexity. Experimental results indicate that the proposed approach improves accuracy and prevents the model from overfitting.

Keywords: hand gesture recognition, IR-UWB, data expansion, multi-input, CNN-LSTM, feature concatenation, multi-class SVM, Optuna

1. INTRODUCTION

Hand Gesture Recognition (HGR) is a very important research area that provides adequate support for several applications such as human-computer interaction and healthcare monitoring [1,2]. A significant effort has been

devoted to gesture recognition using different sensing technologies [3]. Conventional HGR approaches mainly use wearable and optical sensors. These frameworks are highly accurate but represent several drawbacks. Wearable sensors such as gloves require carrying a load of cables that connect the device to a computer while per-

forming the gesture [4]. This makes the system impractical and can cause discomfort for users. In contrast, optical sensors such as cameras do not require any devices attached to the body [5]. However, one prominent concern is the risk of privacy violation when used in personal settings. Furthermore, there arise some situations wherein gesture recognition via cameras is difficult such as sudden lighting changes and the presence of severe occlusions. To overcome the above shortcomings, radar-based sensing systems are proposed [6]. Impulse Radio Ultra-Wide Band (IR-UWB) has recently emerged as one of the most effective and promising non-contact sensors for HGR. It has been deployed in a network fashion for HGR to develop applications such as control car devices [7], wireless keyboards [8,9], and sign language-based communication systems [10,11]. The IR-UWB has the advantage of being remotely operable in a non-intrusive manner. It does not capture any visual images which allow the users to feel unrestrained. Furthermore, the IR-UWB offers an inexpensive and robust system that operates with low power consumption and performs well in both highly lit and dark environments. In addition, it completely avoids the problem of occlusion owing to its high penetration capabilities through obstacles and walls.

The HGR process involves extracting a set of relevant features from the sensor data that best describe a gesture [6], allowing it to be identified with a high recognition rate regardless of the environment in which it is performed or the person performing it [12].

Based on the existing feature extraction techniques, an HGR system can be classified as a traditional or deep model [13]. A traditional model relies on hand-crafted feature extraction, which requires pre-processing of the data to reduce dimensionality and determine appropriate features [14]. Several methods have been investigated for IR-UWB-based HGR, including Multi-Layer Perceptron (MLP) [15], SVM [16], K-Nearest Neighbors (KNN) [17], and K-means [18]. Although these approaches have managed to achieve impressive recognition rates, they are not straightforward, requiring a lot of work for manual feature extraction that heavily depends on human experience and domain knowledge [19]. On the other hand, a deep model eliminates the manual feature extraction phase by replacing it with automatic processing where multi-level features are automatically extracted from raw data, involving less human intervention [7].

In terms of deep models, Convolutional Neural Network (CNN) [20][21] and Recurrent Neural Network (RNN) [22] are the most prominent approaches used for HGR. CNN is considered one of the most efficient deep models for image classification tasks [23]. It acts as a spatial feature extractor and allows one to learn high-level representations in a hierarchical manner using a set of stacked convolutional layers. Several neural network methods have been analyzed, and the results show that CNN is effective for classifying image data generated by IR-UWB for HGR [24-26]. As for RNN, it is

used to analyze sequential data. The most commonly deployed variant of RNN is the Long-Short-Term Memory model (LSTM) [27]. This model is designed with a memory mechanism that uses gates, allowing for exploiting and learning relevant temporal patterns in data. LSTM has been effectively used in various HGR studies with radar [28], and it has also proven to boost performance in terms of classification accuracy when used in a hybrid configuration with CNN [29].

Training deep models from scratch typically requires a large amount of data, which is often not available. Acquiring and annotating remote sensing data can be complex, laborious, and time-consuming, making it challenging to gather the necessary amount of data. As a result, the concept of transfer learning was introduced [30], which involves reusing a previously trained model developed for one task in a new task. However, transfer learning-based algorithms may exhibit unpredictable performance if there is a mismatch between the source and target learning content. Another approach is to transform radar data into different domains, where useful features can be extracted and fused for classification [6,31]. However, this approach requires significant processing and computational resources.

This paper proposes a simple processing approach based on low-level feature extraction to increase the number of samples, as well as a multi-input hybrid model to improve the existing results on an available real-world dataset of dynamic hand gestures acquired using an IR-UWB. Three major points have been addressed in this work. First, the limited amount of data used to train a model. A gesture may not be fully described by a single representation; hence the need to extend the data. The introduction of our data processing using the Sobel filter to extract gradient features significantly filters out unnecessary information while retaining the main features. This process can not only increase the amount of information used to describe a target but also enhance the bottom features by reducing the noise in the data and providing more diversified information. Second, we proposed a three-input CNN-LSTM feature extractor that takes advantage of automatic domain-aware extraction and concatenation of complementary features from the same target to provide more exhaustive spatiotemporal information. Third, the model combines the strength of CNN-LSTM and SVM to improve the recognition accuracy and generalization ability while maintaining a simple architecture with a reasonable number of parameters.

The major contributions of this paper are:

- Using preprocessing steps to extend the amount of data in each class to prevent overfitting.
- Providing a lightweight three-input architecture to process the input data, resulting in considerable improvement in training time.
- Utilising CNN-LSTM layers for automatic spatiotemporal feature learning without any manual en-

gineering or prior domain knowledge.

- Using a multi-class SVM classifier for efficient classification.
- Achieving a high recognition rate and outperforming current state-of-the-art models used for IR-UWB-based hand gesture recognition.

The rest of the paper is organized as follows: Section 2 provides a brief review of recent scholarly works related to HGR-based IR-UWB radar. Section 3 presents a description of the proposed system, including dataset processing and model implementation. Section 4 provides the experimental results and comparative performance of the proposed model. The discussion is presented in Section 5, and Section 6 concludes the paper.

2. RELATED WORKS

2.1. TRADITIONAL MODELS

In the HGR process, a gesture needs to be represented by a suitable set of features. Ghaffar *et al.* [19] used the Histogram of Oriented Gradient (HOG) to extract features from the data of 4 IR-UWB. The resulting features were merged and fed as input to an SVM, resulting in an accuracy of 96% for the classification of 4 gestures. Li *et al.* [17] tested several combinations of Cumulative Distribution Density (CDD) features extracted from IR-UWB spectrograms to train a KNN algorithm, with the highest accuracy achieved being about 82.4%. Khan *et al.* [18] extracted three features, namely the variance of the Probability Density Function (PDF) of the magnitude histogram, Time Of Arrival (TOA) variation, and frequency from the data of an IR-UWB. They used the K-means clustering algorithm to classify 5 gestures, achieving an accuracy of 98%.

2.2. DEEP MODELS

2.2.1. CNN

Ahmed *et al.* [32] proposed a four-layer CNN and tested it on the dataset used in this paper. The task was to automatically extract features from range-time radar spectrograms and classify 12 dynamic hand gestures. A recognition accuracy of 94% was achieved. In another study, Ahmed *et al.* [7] proposed a system to recognize gestures for controlling electronic devices inside a car. The authors used a six-layer CNN to extract features from range-time radar spectrograms converted into grayscale images. The system achieved an accuracy of 96% in recognizing the 5 gestures used in the study. Khan *et al.* [33] employed a five-layer CNN to classify hand gestures based on image trajectory patterns generated from multiple IR-UWB.

2.2.2. LSTM

Noori *et al.* [34] investigated an LSTM architecture tested on the same dataset used in this paper. Their model showed superior performance compared to [32] by reaching an accuracy of 97%. However, it had

847,055 trainable parameters, making it computationally heavy. Park *et al.* [35] proposed a recognition algorithm based on LSTM for the classification of 6 dynamic hand gestures and achieved an accuracy of 90.5%.

2.2.3. CNN-LSTM

Skaria *et al.* [36] implemented a hybrid model that combines CNN and LSTM layers, trained on a 3D tensor of stacked range-Doppler frames. The CNN-LSTM achieved a high accuracy of 96.15%. In another work by the same authors, two sensors were investigated for robust gesture classification, namely an IR-UWB and a thermal sensor [37]. CNN-LSTM layers were employed on both radar and thermal signals, achieving an accuracy of 99% for 14 hand gestures.

3. MATERIALS AND METHOD

3.1. DATASET

This study aims to improve the existing results on the public UWB Gestures dataset proposed by Ahmed *et al.* [32]. The dataset consists of 12 classes of dynamic hand gestures, namely left-right (LR), right-left (RL), up-down(UD), down-up (DU), diag-LR-UD, diag-LR-DU, diag-RL-UD, diag-RL-DU, clockwise, counter-clockwise, push-in, and empty gestures. Each gesture is performed 100 times by 8 volunteers and acquired using three XeThru X4 IR-UWBs (Fig. 1). In order to compare the performance of our three-input CNN-LSTM-SVM against other models using the same dataset, we followed the same procedure and used only the left radar data, which consists of 9,600 range-time images.

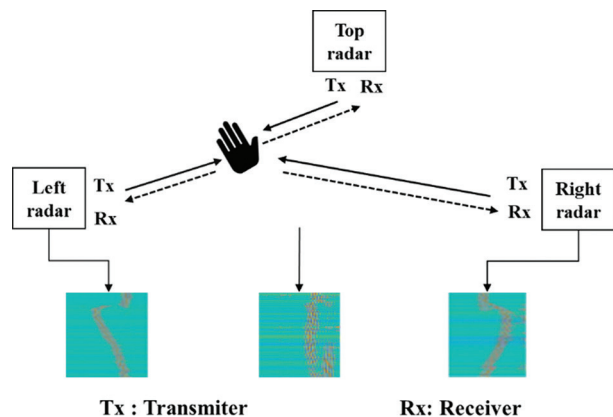


Fig. 1. Collection of the UWB-Gestures dataset using three UWB radars.

3.2. DATA PROCESSING

A major problem in training deep models from scratch is the huge amount of data required. Using a small dataset usually results in overfitting and decreased model performance. To achieve high generalization capability, we propose extending the dataset and generating low-level image features to use as an in-

dependent input for the model. These features include the image gradient in the x-direction, y-direction, and both x and y-directions generated by the Sobel filter. First, the samples are resized to 75x75 to decrease the computational cost, then converted to binary images by applying thresholding. Next, the Sobel filter is applied in the horizontal and vertical directions to calculate the gradient in the x-direction (Dx) and y-direction (Dy) for each sample. Finally, the gradient images for both x and y-directions are generated by taking the square root of the sum of Dx and Dy for each sample. By doing this, the size of the dataset is tripled, and a total of 25,200 images are generated without creating duplicates (Fig. 2).

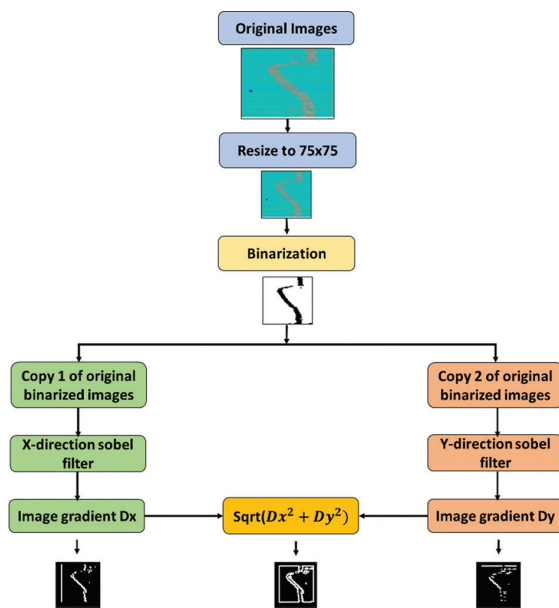


Fig. 2. Flow graph of data processing.

3.3. METHOD

The proposed approach is designed to classify dynamic hand gestures, which are formed by a consecutive sequence of poses where their characteristics vary over time. Therefore, the recognition process of dynamic gestures takes into account both spatial and temporal information. To enable the analysis of spatial and temporal features of gestures, we propose using a three-input CNN in conjunction with an LSTM to process the input data and perform feature extraction. First, the three-input CNN is used to extract the spatial features corresponding to the low-level representation of the images. Each representation is processed separately in a CNN branch. To fully represent the gesture, the output features of the three CNN branches are merged. The concatenated features are reshaped and provided as input to the LSTM for temporal feature extraction. The LSTM captures and memorizes how the features extracted by the CNN layers change over time. The output of the LSTM is put into vector form and fed into the multiclass SVM. The SVM acts as the final classifier of the architecture and gives the prediction result (Fig. 3).

The noteworthy characteristics of the three-input CNN-LSTM-SVM are:

- Increasing the amount of input data to prevent overfitting and improve the generalization performance of the model.
- Adopting a more efficient feature extractor.
- Using both spatial and temporal information to describe a gesture.
- Providing the final classifier with more information to make a decision.

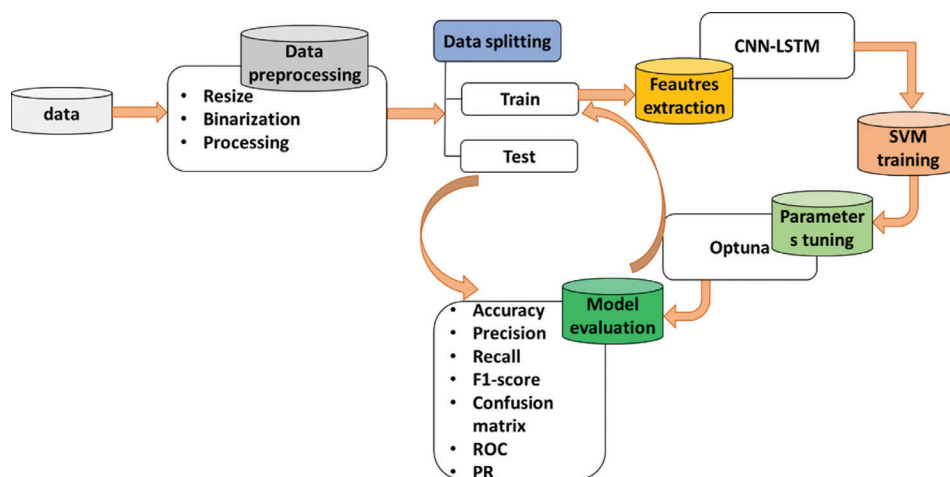


Fig. 3. The overall structure of the proposed hand gesture recognition system

3.3.1. Learning spatiotemporal features

The combination of both spatial and temporal features is a requirement for dynamic gesture classification. To achieve this, two models are combined: the three-input CNN and the LSTM.

The three-input CNN is used for spatial feature extraction. It consists of three branches with similar layer configurations. Each branch consists of an input layer of size 75x75 and three convolutional layers with 16, 32, and 64 filters of size 4x4, respectively, which reduces the number of the training parameters. Additionally, 2x2 strides

are used to further reduce the computational cost. Each branch takes a different image representation as input. The images are processed separately by performing multiple convolution operations to extract spatial features. The convolution operation is expressed as follows:

$$F(i, j) = (I * K)(i, j) = \sum \sum I(i + m, j + n)K(m, n) \quad (1)$$

where I is the input image, i and j are the height and width respectively, K is the 2D convolutional filter of size $m \times n$, and F is the output 2D feature map. To increase nonlinearity in feature maps, a Maxout layer is inserted. It is mathematically expressed as follows:

$$f(x) = \max_{j \in [1, k]} Z_{ij} \quad (2)$$

$$Z_{ij} = x^T W_{ij} + b_{ij}.$$

with x is the input variables, W the weight, and b the bias.

According to the article published by Goodfellow et al. [38], the Maxout activation function has demonstrated its effectiveness for training with Dropout, as well as its robustness for image classification [39]. Each feature map is dimensionally reduced using a Maxpooling layer with a pool size of 2×2 to preserve the most relevant features identified and avoid unnecessary computations. Finally, a Dropout layer with a value of 25% is inserted. The three CNN branches operate in parallel, and their outputs are combined by late fusion for further processing by the LSTM.

The LSTM layer is composed of 150 units. The structure of an LSTM unit consists of input, output, and forget gates that control the learning process, as shown in Fig. 4. These gates are adjusted using the activation sigmoid function. To avoid overfitting in the recurrent layer, the recurrent dropout is set to 0.2.

3.3.2. Multi-class SVM for gesture classification

The spatiotemporal features automatically generated by the CNN-LSTM are fed into the SVM module for training and testing on the hand gesture dataset (Fig. 5). Since the dataset we are using consists of $M = 12$ classes, we have implemented a multi-class SVM algorithm based on the combination of a set of M binary SVMs. We decomposed the multiclass problem into 64 filters of size 4×4 , respectively, which reduces the number of the training parameters. Additionally, 2×2 strides are used to further reduce the computational cost. Each branch takes a different image representation as several bi-class problems and adopted "the one against all" strategy, where each binary classifier is trained on the samples of a selected class against all other classes.

This means that the samples on which a classifier is trained are labeled as positive, and all the rest are labeled as negative. In the evaluation phase, a test sample is labeled as belonging to a class according to the maximum score among the 12 classifiers.

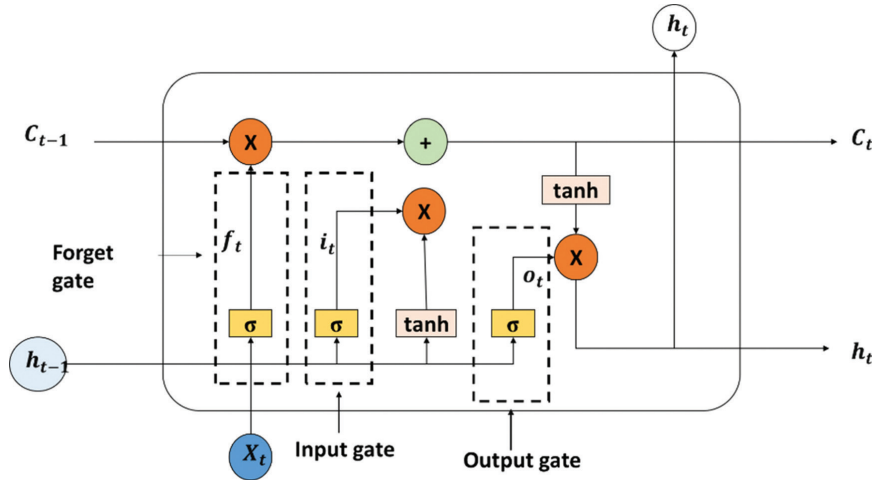


Fig. 4. LSTM structure

3.3.3. Evaluation method

To evaluate the performance of our proposed model, we use the following metrics: accuracy, precision, recall, and F1-score. These metrics are calculated based on the number of true positives (T_p), true negatives (T_N), false positives (F_p), and false negatives (F_N) using the following equations:

$$Accuracy = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (3)$$

$$Precision = \frac{T_p}{T_p + F_p} \quad (4)$$

$$Recall = \frac{T_p}{T_p + F_N} \quad (5)$$

$$F1-score = \frac{2T_p}{2T_p + F_p + F_N} \quad (6)$$

3.3.4. Implementation details

The three-input CNN-LSTM-SVM model is implemented in Python using the Keras framework with Tensorflow on a machine running an environment with an Intel (R) Core (TM) i5 2.40 GHz CPU, 16GBs of RAM, 1TB of hard disk, and Windows 10. The dataset samples are

randomly split into 80% for training and 20% for testing using the `train_test_split` function included in the `sklearn` python package. The random seed parameter is also used to ensure that the test samples are the same in each performed experiment. During the training process of the three-input CNN-LSTM, the Adam optimizer is used with a learning rate set to 0.001, a batch size of 16, and 25 epochs. The input labels are provided as integers rather than vectors to save time in memory as well as computations. Therefore, we use the sparse categorical cross-entropy loss function. The resulting spatiotemporal feature vectors from the three-input CNN-LSTM are

fed to the multi-class SVM classifier for training and testing. The learning process of the multi-class SVM classifier is performed using the Optuna framework [40]. During the learning process, the hyperparameters tuning, including the penalty coefficient C , kernel function, slack variable (degree), and gamma parameter, is achieved based on several repeated trials using k cross-validation with $k = 5$. Optuna is an open-source optimization software available as a library in Python that is easy to implement and offers an integrated dashboard to visualize optimization histories and results. The source code of our model is available on GitHub [41].

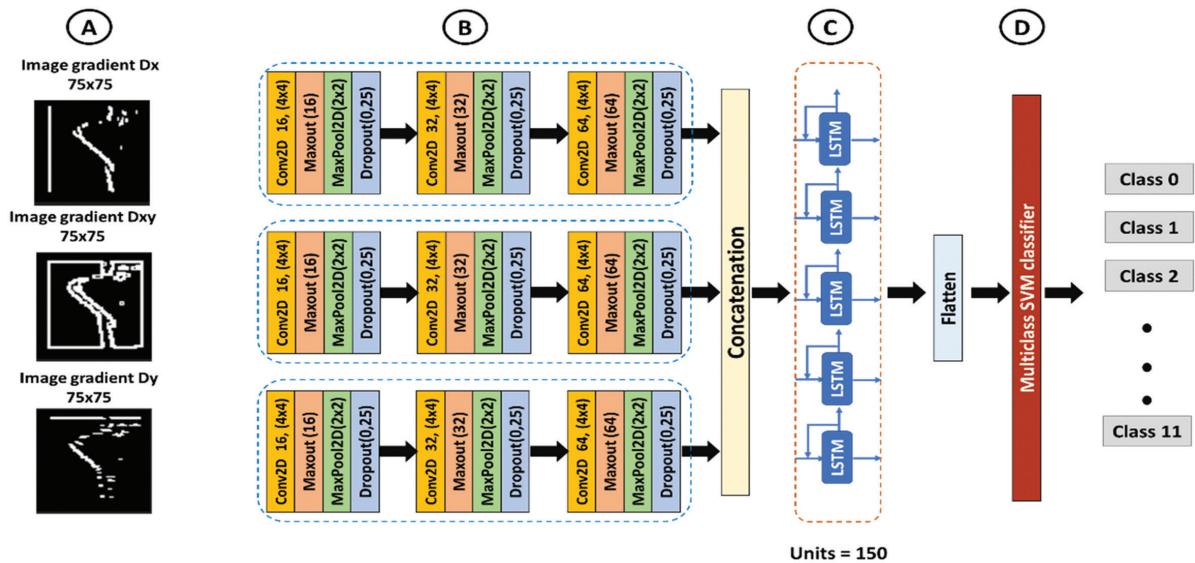


Fig. 5. Three-input CNN-LSTM-SVM model:(A) Input layer, (B) CNN layers for spatial features extraction, (C) LSTM layer for learning temporal features, (D) Classification/output layer.

4. EXPERIMENTAL RESULTS

4.1. TRAINING PROCESS

The training process is divided into three main phases:

- The CNN-LSTM is trained with a SoftMax classifier for the extraction of spatiotemporal features.
- The SoftMax is replaced by the Multiclass SVM, which is fed with the feature vectors, trained, and optimized for 100 trials.
- The Multiclass SVM is then trained for a second time with the optimal hyperparameters.

The finetuned Multiclass SVM achieved a training accuracy of 99.62% (Fig. 6). The influence of different hyperparameters values on the model's performance is presented in Fig. 7.

4.2. EVALUATION PROCESS

The finetuned three-input CNN-LSTM-SVM model is evaluated on the test set. The confusion matrix and the classification report obtained from the test data are depicted below (Fig. 8). Additionally, Receiver Operating Characteristics (ROC) and Precision- Recall (PR) curves are plotted to compare the overall performance (Fig. 9).

4.3. COMPARISON

4.3.1. Verification of data processing

To demonstrate the effectiveness of the proposed data processing approach, the first experiment is divided into two main parts. In the first part, the three-input CNN-LSTM-SVM model is trained using raw images (original images without processing), where the same sample is simultaneously provided to all three CNN branches. In the second part, the model is trained using extended images, where each CNN branch is fed with a different representation of low-level features. The results are shown in Table 1.

Table 1. Comparative classification performance on Raw/Processed images.

Metrics	Raw images	Processed images
Train accuracy	98.34%	99.62%
Test accuracy	92.49%	98.27%
Precision	92.77%	98.30%
Recall	92.40%	98.29%
F1-score	92.44%	98.27%

Optimization History Plot

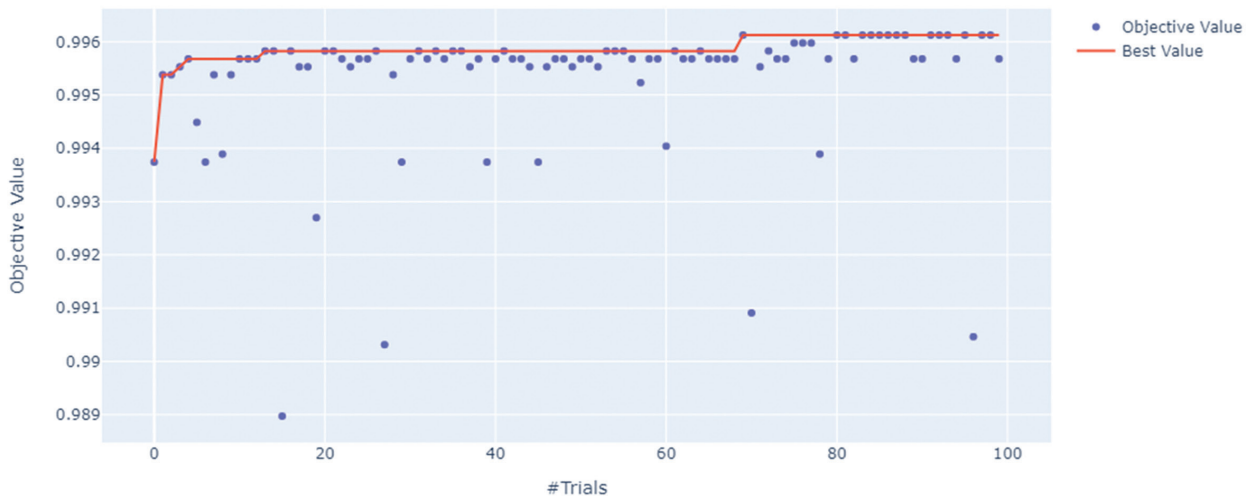


Fig. 6. Optimization history.

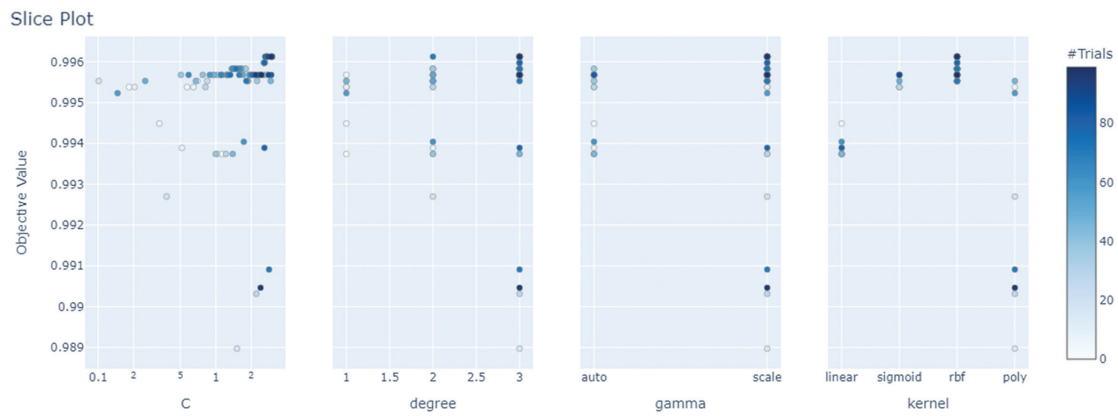


Fig. 7. Slice plot.

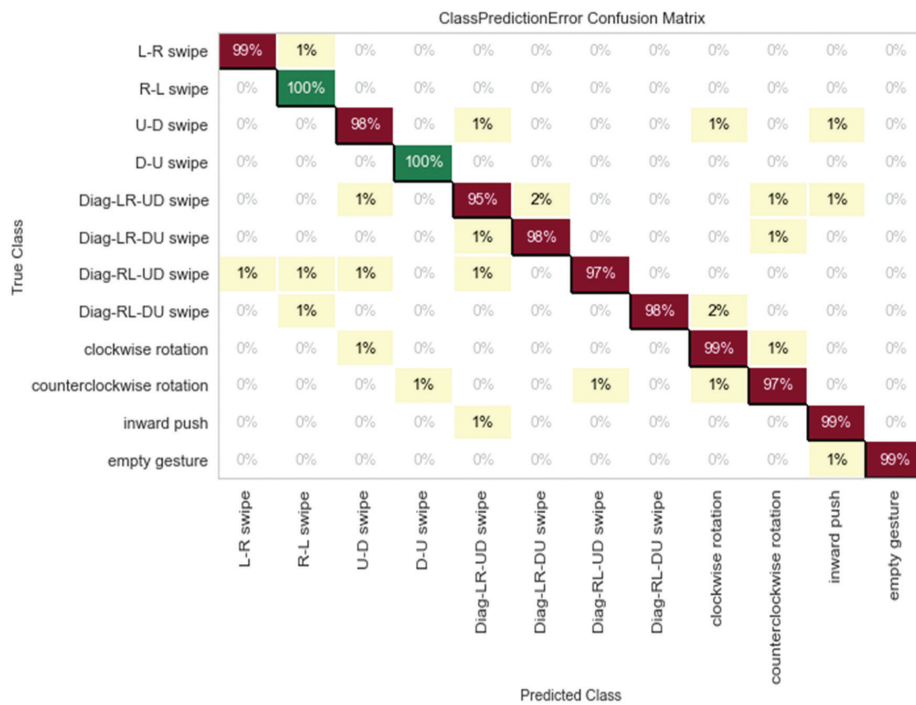


Fig. 8. (a) Confusion matrix

ClassPredictionError Classification Report

empty gesture	1.000	0.986	0.993	139
inward push	0.972	0.993	0.982	138
counterclockwise rotation	0.974	0.974	0.974	154
clockwise rotation	0.966	0.986	0.976	146
Diag-RL-DU swipe	1.000	0.976	0.988	126
Diag-RL-UD swipe	0.994	0.975	0.984	158
Diag-LR-DU swipe	0.981	0.981	0.981	157
Diag-LR-UD swipe	0.969	0.955	0.962	132
D-U swipe	0.993	1.000	0.996	142
U-D swipe	0.977	0.977	0.977	133
R-L swipe	0.976	1.000	0.988	124
L-R swipe	0.992	0.992	0.992	129

precision recall f1 support




Fig. 8. (b) Classification report

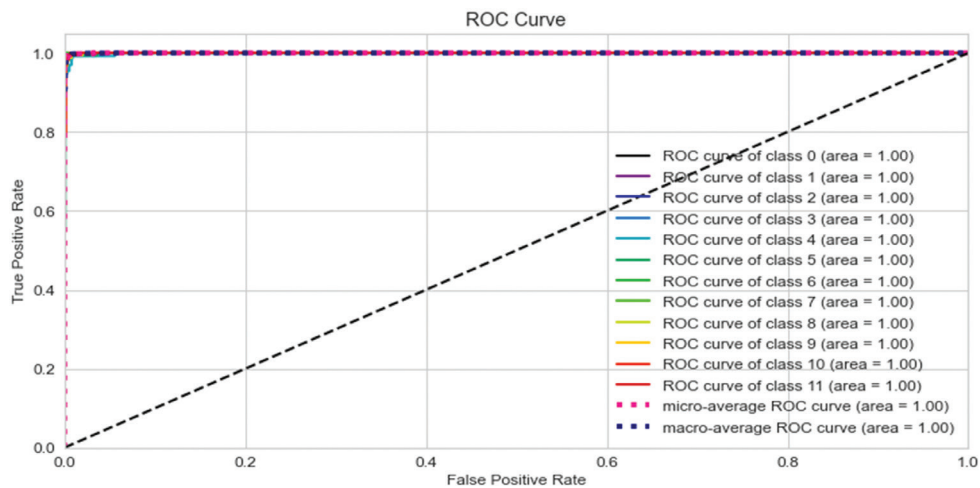


Fig. 9. (a) ROC curve

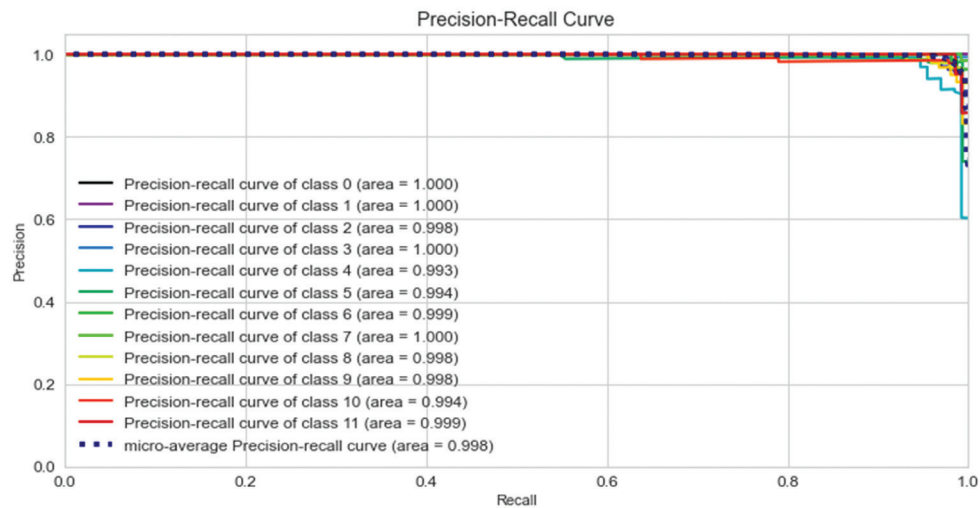


Fig. 9. (b) PR curve

4.3.2. Verification of the model structure

The second experiment aims to demonstrate the superiority of using multiple branches for parallel data processing over a single branch for sequential data processing when using different feature representations. We fed a single-input CNN-LSTM-SVM with the extended images and compared its performance to the results obtained from the three-input CNN-LSTM-SVM experiment mentioned in subsection 4.3.1. Note that the single-input CNN-LSTM-SVM used in this experiment consists of the same layer and parameter configuration as a single CNN branch from the three-input CNN-LSTM-SVM. The results are presented in Table 2.

Table 2. Comparative classification performance of single/three-input CNN-LSTM-SVM.

Metrics	single-input CNN-LSTM-SVM	three-input CNN-LSTM-SVM
Train accuracy	96.30%	99.62%
Test accuracy	93.09%	98.27%
Precision	93.24%	98.30%
Recall	93.10%	98.29%
F1-score	93.44%	98.27%

4.3.3 Comparison to state-of-the-art approaches

The third experiment aims to compare the classification performance of the three-input CNN-LSTM-SVM with other models, including three-input CNN-SoftMax and three-input CNN-LSTM-SoftMax models. This experiment is performed to demonstrate the impact of using spatial features only versus using spatiotemporal features on the recognition rate. Moreover, we aim to find the optimal classifier for the model by comparing SoftMax and SVM. Furthermore, our proposed model results can be directly compared with those of Ahmed et al. [32] and Noori et al. [34], as they also used the same dataset. The results are shown in Table 3.

Table 3. Comparative classification performance of three-input CNN-LSTM-SVM with state of art methods

Reference	Model	Accuracy	N° of parameters
Ahmed et al. [32]	CNN	94%	-
	Three input-CNN-SoftMax	95.41%	126300
Noori et al. [34]	LSTM	97%	847055
	Three input-CNN-LSTM-SoftMax	97.20%	331596
Our approach	Three input-CNN-LSTM-SVM	98.27%	332578

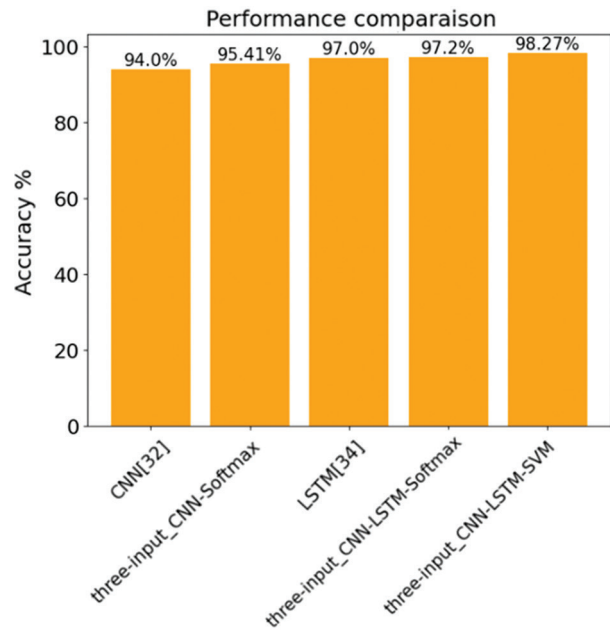


Fig. 10. Comparison of proposed three-input CNN-LSTM-SVM model accuracy with state of art methods.

5. DISCUSSION

This work presents an end-to-end hybrid model for classifying dynamic hand gestures using IR-UWB data. We used a three-input CNN-LSTM for automatically learning spatiotemporal features, combined with a multiclass SVM classifier. The performance of the proposed model was assessed using various evaluations, and the results of the confusion matrix and classification report are depicted in Fig. 8. They show that the model achieved an accuracy of 98.27% and identified half of the classes with an accuracy above 98%, a precision of 98.30%, a recall of 98.29%, and an F1-score of 98.27%. Moreover, the ROC curve in Fig. 9 indicates that the proposed model based on combined features produced excellent results, with a true positive rate produced for each gesture class.

From Table 1, it is evident that the three-input CNN-LSTM-SVM model is overfitting when duplicating inputs using raw images, as it achieved a training accuracy of 98.34% but a lower test accuracy of 92.42%. This is due to the non-representative dataset, which does not provide enough information for the model to generalize well. To address this issue and prevent the model from overfitting, we used the extended dataset that includes image gradients in the x-direction, y-direction, and both x and y-directions. As shown in Table 1, the three-input CNN-LSTM-SVM model achieves better gesture recognition when different low-level representations are provided as inputs, leading to an increase in accuracy of about 6% compared to using duplicate original images. The model achieved a training and test accuracy of 99.62% and 98.27%, respectively, with

much lower overfitting and better generalization ability. This is because using multiple representations for the same gesture enables the extraction and merging of more features, providing the final classifier with more information to make a decision. These results highlight the importance of using different low-level representations as well as a multiple inputs model in cases where there are insufficient representative features, and indicate that unnecessary duplication of inputs does not significantly improve performance but increases computational complexity.

The superiority of our method is more apparent when we compare the performance of the single-input and three-input models. The results in Table 2 show a significant difference in the model's performance when processing the data sequentially and independently in parallel. The single-input CNN-LSTM-SVM achieved an accuracy of 93.09%, while the three-input CNN-LSTM-SVM achieved an accuracy of 98.27%. Compared to the single-input CNN-LSTM-SVM, our proposed three-input CNN-LSTM-SVM architecture not only leverages the strength of multiple low-level representations to extract complementary features from the same target but also introduces the concept of feature concatenation into the architecture to achieve more holistic representations. The separate processing of each data representation allows the extraction and preservation of its features without altering them with other data representations, enabling the model to learn distinct, discriminative, and complementary features effectively. We can conclude that in limited data conditions, extending raw data to different feature representations and then providing them as inputs should be applied to separate branches to achieve higher accuracy. We also hypothesize that the reason for the low performance of the single-input CNN-LSTM-SVM model is due to dissimilar features, where the same gesture is represented three times differently, leading to confusion for the model and making it error-prone. We can reasonably conclude from this experiment that through the effective concatenation of multiple feature information, the prediction is made based on the full use of target features, which improves the recognition accuracy.

On the other hand, as shown in Table 3, the proposed three-input CNN-LSTM-SVM outperforms the most recent research works (Fig. 10) [32,34]. Compared to the CNN proposed by Ahmed et al. [32], the three-input CNN-SoftMax achieved an average increase in accuracy of 1.41%. This result suggests that the three-input CNN-SoftMax can capture more spatial context information for classification by learning the details when a large number of training samples are provided. Although data expansion contributes to the model, the accuracy is still limited by the lack of additional information. While the CNN structure can learn higher-level features, it ignores the temporal dependencies on the features, meaning that the inputs and outputs are independent, leading to limited recognition performance.

We can observe from Table 3 that adding LSTM units can boost classification performance. The combination of three-input CNN-LSTM-SoftMax showed its effectiveness by achieving an accuracy of 97.20%, outperforming the CNN proposed by Ahmed et al. [32] by 3.20% and the three-input CNN-SoftMax by 1.79%. This enhancement refers to convolution, concatenation operations, and the sophisticated structure of LSTM, that maintain the spatial and temporal relationships. The LSTM is cascaded to learn and integrate temporal features, which can provide additional information and improve classification performance. The LSTM helps capture and memorize how the features extracted by the CNN layers change over time. Combining the strengths of CNN and LSTM provides the benefits of both spatial and temporal learning, which is very effective in improving the recognition rate of dynamic hand gestures. Moreover, the three-input CNN-LSTM-SoftMax model provided comparable performance to previous work by Noori et al. [34]. However, the three-input CNN-LSTM-SoftMax model achieved 97.20% accuracy and had fewer trainable parameters (331,596) than the model proposed by Noori et al. [34], which was maintained with 847,055 parameters and achieved 97% accuracy. To select the optimal classifier for our model, the three-input CNN-LSTM was trained with a SoftMax layer and an SVM classifier. The results reported in Table 3 show a 1.07% increase in accuracy using the SVM as the final classifier. This gain is mainly due to the use of the various optimal hyperparameters selected using the Optuna framework. Therefore, the generalization ability of SVM is superior to that of SoftMax.

6. CONCLUSION

A major concern when training deep learning models is the requirement for a large amount of data to achieve sufficient robustness. Otherwise, with limited data, models are prone to overfitting. To enrich the training and testing samples in radar-based HGR, this paper proposes a simple and efficient method to extend radar spectrograms. Using different low-level feature representations as input, processed on separate branches, helps the model learn and merge more information about the target, resulting in increased accuracy. Combining CNN and LSTM layers to take into account both spatial and temporal features improve recognition accuracy. Finally, switching from SoftMax to SVM appears to be beneficial for generalization ability.

We believe that the number of samples, as well as their representation in the dataset, are critical factors in developing a robust model that provides high classification predictions. Future work in this research aims to introduce more data processing techniques to generate additional samples to enhance the model's performance in terms of similar evaluation parameters. Additionally, we plan to further reduce the model's complexity by modifying the layer configuration.

7. REFERENCES

- [1] L. Guo, Z. Lu, L. Yao, "Human-Machine Interaction Sensing Technology Based on Hand Gesture Recognition: A Review", *IEEE Transactions on Human-Machine Systems*, Vol. 51, No. 4, 2021, pp. 300-309.
- [2] B. Van Amsterdam, M. J. Clarkson, D. Stoyanov, "Gesture Recognition in Robotic Surgery: A Review", *IEEE Transactions on Biomedical Engineering*, Vol. 68, No. 6, 2021, pp. 2021-2035.
- [3] S. Wu, Z. Li, S. Li, Q. Liu, W. Wu, "An overview of gesture recognition", *Proceedings of the International Conference on Computer Application and Information Security*, Wuhan, China, 23-24 December 2022, pp. 600-606.
- [4] M. Pan, Y. Tang, H. Li, "State-of-the-Art in Data Gloves: A Review of Hardware, Algorithms, and Applications", *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, 2023, pp. 1-15.
- [5] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human-computer interaction", *Institute of Engineering and Technology Computer Vision*, Vol. 12, No. 1, 2018, pp. 3-15.
- [6] S. Ahmed, K. D. Kallu, S. H. Cho, "Hand gestures recognition using radar sensors for human computer-interaction: A review", *Remote Sensing*, Vol. 13, No. 3, 2021, p. 527.
- [7] S. Ahmed, F. Khan, A. Ghaffar, F. Hussain, S. H. Cho, "Finger-counting-based gesture recognition within cars using impulse radar with convolutional neural network", *Sensors*, Vol. 19, No. 6, 2019, pp. 1429.
- [8] S. K. Leem, F. Khan, S. H. Cho, "Detecting Mid-Air Gestures for Digit Writing with Radio Sensors and a CNN", *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 4, 2020, pp. 1066-1081.
- [9] N. Hendy, H. M. Fayek, A. Al-Hourani, "Deep Learning Approaches for Air-Writing Using Single UWB Radar", *IEEE Sensors Journal*, Vol. 22, No. 12, 2022, pp. 11989-12001.
- [10] H. Hameed, M. Usman, M. Z. Khan, A. Hussain, H. Abbas, M.A. Imran, Q. H Abbasi, "Privacy-Preserving British Sign Language Recognition Using Deep Learning", *Proceedings of the 44th International Conference of the IEEE Engineering in Medicine & Biology Society*, Glasgow, Scotland, UK, 11-15 July 2022, pp. 4316-4319.
- [11] Y. Yang, J. Li, B. Li, Y. Zhang, "MDHandNet: a light-weight deep neural network for hand gesture/sign language recognition based on micro-doppler images", *World Wide Web*, Vol. 25, No. 5, 2022, p. 1951-1969.
- [12] D. Sarma, M. K. Bhuyan, "Methods, Databases and Recent Advancement of Vision-Based Hand Gesture Recognition for HCI Systems: A Review", *SN Computer Science*, Vol. 2, No. 6, 2021, p. 436.
- [13] Y. Shi, Y. Li, X. Fu, K. Miao, Q. Miao, "Review of dynamic gesture recognition", *Virtual Reality & Intelligent Hardware*, Vol. 3, No. 3, 2021, pp. 183-206.
- [14] J. Park, S. H. Cho, "IR-UWB radar sensor for human gesture recognition by using machine learning", *Proceedings of the 18th International Conference on High Performance Computing and Communications, 14th International Conference on Smart City, 2nd International Conference on Data Science and Systems*, Sydney, NSW, Australia, 12-14 December 2016, pp. 1246-1249.
- [15] K. Faheem, L. Seong Kyu, S. H. Cho, "Algorithm for fingers counting gestures using IR- UWB radar sensor", *Proceedings of the International IEEE Sensors Applications Symposium*, Seoul, Korea, 12-14 March 2018, pp.144-146.
- [16] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, T. W. Kim, "A hand gesture recognition sensor using reflected impulses", *IEEE Sensors Journal*, Vol. 17, No. 10, 2017, pp. 2975-2976.
- [17] B. Li, J. Yang, Y. Yang, C. Li, Y. Zhang, "Sign Language/ Gesture Recognition Based on Cumulative Distribution Density Features Using UWB Radar", *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, 2021, pp. 1-13.
- [18] F. Khan, S. Leem, S. H. Cho, "Hand-based gesture recognition for vehicular applications using IR-UWB radar", *Sensors*, Vol. 17, No. 4, 2017, p. 883.
- [19] A. Ghaffar, F. Khan, S. H. Cho, "Hand Pointing Gestures Based Digital Menu Board Implementation Using IR-UWB Transceivers", *IEEE Access*, Vol. 7, 2019, pp. 58148-58157.
- [20] L. Yao, W. Xin, S. Baodai, M. Zhu, "Hand Gesture Recognition Using IR-UWB Radar with ShuffleNet V2", *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence*, San-ya, China, 14-16 January 2021, pp. 126-131.

- [21] A. Bhavana, K. S. R. Kumar, M. D. Praveen, "Deep Neural Network based Sign Language Detection", Proceedings of the 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 1-3 December 2022, pp. 1474-1479.
- [22] J. J. Ojeda-Castelo, M. D. L. M. Capobianco-Uriarte, J. A. Piedra-Fernandez, R. Ayala, "A Survey on Intelligent Gesture Recognition Techniques", IEEE Access, Vol. 10, 2022, pp. 87135-87156.
- [23] Z. Li, W. Yang, S. Peng, F. Liu, "A Survey of Convolutional Neural Networks: Analysis, Applications and Prospects", IEEE Transactions on Neural Networks and Learning Systems, Vol. 33, No. 12, 2021, pp. 6999-7019.
- [24] S. Ahmed, S. H. Cho, "Hand gesture recognition using an IR-UWB radar with an inception module-based classifier", Sensors, Vol. 20, No. 2, 2020, p. 564.
- [25] G. Park, V. K. Chandrasegar, J. Park, J. Koh, "Increasing Accuracy of Hand Gesture Recognition using Convolutional Neural Network", Proceedings of the International Conference on Artificial Intelligence in Information and Communication, Jeju Island, Korea, 21-24 February 2022, pp. 251-255.
- [26] G. Park, V. K. Chandrasegar, J. Koh, "Accuracy Enhancement of Hand Gesture Recognition using CNN", IEEE Access, Vol. 11, 2023, pp. 26496-26501.
- [27] Y. Yu, X. Si, C. Hu, J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures", Neural Computation, Vol. 31, No. 7, 2021, pp. 1235-1270.
- [28] H. Liu, Z. Liu, "A Multi-Modal Dynamic Hand Gesture Recognition Based on Radar-Vision Fusion", IEEE Transactions on Instrumentation and Measurement, Vol. 72, 2023, pp. 1-15.
- [29] S. Skaria, A. Al-Hourani, Da Huang, "Radar-Thermal Sensor Fusion Methods for Deep Learning Hand Gesture Recognition", Proceedings of the International Conference of IEEE Sensors, Sydney, Australia, 31 October - 3 November 2021, pp. 1-4.
- [30] L. O. Fhager, S. Heunisch, H. Dahlberg, A. Evertsson, L. E. Wernersson, "Pulsed Millimeter Wave Radar for Hand Gesture Sensing and Classification", IEEE Sensors Letters, Vol. 3, No. 12, 2019, pp. 1-4.
- [31] S. Z. Gurbuz, M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring", IEEE Signal Processing Magazine, Vol. 36, No. 4, 2019, pp. 16-28.
- [32] S. Ahmed, D. Wang, J. Park, S. H. Cho, "UWB-gestures, a public dataset of dynamic hand gestures acquired using impulse radar sensors", Scientific Data, Vol. 8, No. 1, 2021, pp. 1-9.
- [33] F. Khan, S. K. Leem, S. H. Cho, "In-Air Continuous Writing Using UWB Impulse Radar Sensors", IEEE Access, Vol. 8, 2020, pp. 99302-99311.
- [34] F. M. Noori, M. Z. Uddin, J. Torresen, "Ultra-Wideband Radar-Based Activity Recognition Using Deep Learning", IEEE Access, Vol. 9, 2021, pp. 138132-138143.
- [35] J. Park, J. Jang, G. Lee, H. Koh, C. Kim, T. W. Kim, "A Time Domain Artificial Intelligence Radar System Using 33-GHz Direct Sampling for Hand Gesture Recognition", IEEE Journal of Solid-State Circuit, Vol. 55, No. 4, 2020, pp. 879-888.
- [36] S. Skaria, A. Al-Hourani, R. J. Evans, "Deep-learning methods for hand-gesture recognition using ultra-wideband radar", IEEE Access, Vol. 8, 2020, pp. 203580-203590.
- [37] S. Skaria, D. Huang, A. Al-Hourani, R. J. Evans, M. Lech, "Deep-Learning for Hand-Gesture Recognition with Simultaneous Thermal and Radar Sensors", Proceedings of the International Conference of IEEE Sensors Conference, Rotterdam, Netherlands, 25-28 October 2020, pp. 1-4.
- [38] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, "Maxout networks", Proceedings of the International Conference on Machine Learning Research, Atlanta, Georgia, USA, 17-19 June 2013, pp. 1319-1327.
- [39] R. Rajeswari, M. Prabhakar, G. Padmapriya, B. S. Kumar, "Blood vessel detection using enhanced DeepJoint fuzzy clustering algorithm with deep Maxout network for glaucoma detection", Concurrency and Computation: Practice and Experience, Vol. 35, No. 6, 2023, p. 1.
- [40] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework", Proceedings of the 25th International Conference on Knowledge discovery & Data mining, Anchorage, AK, USA, 4-8 August 2019, pp. 2623-2631.
- [41] GitHub, three_input_CNN_LSTM_SVM, https://github.com/souhila1998/HGR_CNN-LSTM-SVM (accessed: 2023)

Performance Analysis of a new Filter and Wrapper Sequence for the Survivability Prediction of Breast Cancer Patients

Original Scientific Paper

E. Jenifer Sweetlin

Research Scholar,
Centre for Information Technology and Engineering,
Manonmaniam Sundaranar University, Tirunelveli, India
jsweetlin@gmail.com

S. Saudia

Assistant Professor,
Centre for Information Technology and Engineering,
Manonmaniam Sundaranar University, Tirunelveli, India
saudiasubash@msuniv.ac.in

Abstract – Feature selection is an essential preprocessing step for removing redundant or irrelevant features from multidimensional data to improve predictive performance. Currently, medical clinical datasets are increasingly large and multidimensional and not every feature helps in the necessary predictions. So, feature selection techniques are used to determine relevant feature set that can improve the performance of a learning algorithm. This study presents a performance analysis of a new filter and wrapper sequence involving the intersection of filter methods, Mutual Information and Chi-Square followed by one of the wrapper methods: Sequential Forward Selection and Sequential Backward Selection to obtain a more informative feature set for improved prediction of the survivability of breast cancer patients from the clinical breast cancer dataset, SEER. The improvement in performance due to this filter and wrapper sequence in terms of Accuracy, False Positive Rate, False Negative Rate and Area under the Receiver Operating Characteristics curve is tested using the Machine learning algorithms: Logistic Regression, K-Nearest Neighbour, Decision Tree, Random Forest, Support Vector Machine and Multilayer Perceptron. The performance analysis supports the Sequential Backward Selection of the new filter and wrapper sequence over Sequential Forward Selection for the SEER dataset.

Keywords: accuracy, filter-wrapper, Sequential forward selection, Sequential backward selection

1. INTRODUCTION

Breast cancer is one of the most serious medical reasons associated with death of women on earth. The disease is caused by many factors such as age, obesity, alcoholism, lack of physical activity, menopausal status and family history of breast cancer [1]. Data Mining and Machine Learning (ML) techniques have been used to develop various breast cancer prediction models [2]. The digital form of clinical breast cancer datasets is available in huge volumes and are multidimensional. This multidimensional dataset contains many independent features with more missing values, and some of the features are irrelevant for analysis. Applying these datasets to ML based classifiers drastically reduces accuracy. So, finding the relevant and optimal feature set combination is more important for enhancing and improving the accuracy of ML based classifiers. The

training phases of the ML model design includes data preprocessing, feature selection and feature extraction stages [3-5]. Medical datasets are mostly imbalanced, so to reduce the effect of skewed class distribution in the model, studies have focused on data balancing methods [6]. To build a more effective ML classifier, feature selection techniques are used to filter out and find more optimal features from multidimensional datasets with irrelevant independent features [7]. Feature selection focuses on selecting significant independent features to improve the ability of classifiers to discriminate between classes. Furthermore, feature selection reduces feature dimensionality and computational complexity during the training phase of an ML algorithm [8].

The feature selection algorithms are categorized as filter, wrapper and embedded methods depending on how they combine feature selection sequences while determining informative independent features. Ranking

technique is the principal criterion of filter methods which use rank ordering for feature selection based on a statistical score. The method filters irrelevant features which degrade the relationship between independent and dependent features, thereby selecting the highly ranked independent features to be applied to the training phase of an ML algorithm. These filter approaches are independent to any ML algorithm, computationally quick and scalable. Some feature ranking based filter techniques are the Mutual Information (MI), Pearson Correlation Coefficient (PCC) and Chi-Square (CS) methods [9]. Compared to filter methods, the wrapper methods show better performance because the independent feature selection mainly depends on a classification algorithm. Wrapper methods determine the quality of different subsets of independent features which are more suited for the classifier. But if the dimensionality of the dataset is large, the wrapper methods are very expensive in terms of time and computational speed since each feature set considered must be evaluated with the ML classifiers used [10]. In embedded methods [11], both filter and wrapper methods are used for feature selection and a classifier is used to evaluate the quality of the selected subset of independent features.

Liou et al. [12] used a Genetic Algorithm (GA) based approach along with an Artificial Neural Network (ANN), Decision Tree (DT) and Logistic Regression (LR) on the Wisconsin Breast Cancer Database-Original (WBCD) dataset for predicting breast cancer. The feature selection step was carried out based on Information Gain (IG). It is indicated in the paper that the GA model yielded better results while classifying the breast cancer data with an accuracy of 98.78%. Saygili [13] studied and compared the diagnosis of cancer using six ML methods such as Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), Naive Bayes (NB), DT, Random Forest (RF) and Multilayer Perceptron (MLP) on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Gain Ratio (GR) was used as a feature selection technique. The results showed that the RF performed better with an accuracy of 98.77%. Omondigbe et al. [14] proposed an automated method for diagnosing breast cancer on the WDBC dataset using SVM, ANN and NB with Correlation Based Filters, Recursive Feature Elimination (RFE), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in the preprocessing stage. The proposed SVM-LDA and ANN-LDA approaches achieved an accuracy of 98.82%. Islam et al. [15] compared five supervised ML techniques, namely SVM, K-NN, RF, ANN and LR on the WBCD dataset. PCC was used to identify the relationship between the attributes. The results showed that ANNs performed well with the highest accuracy of 98.57%. Alickovic et al. [16] used two datasets, WDBC and WBCD to construct an automated breast cancer diagnosis system to select significant features from the datasets; GA was used as a feature selection technique. The authors compared the performance of classifiers with GA feature selec-

tion and without GA feature selection on data mining algorithms such as DT, LR, Bayesian Network (BN), RF, Radial Basis Function Networks (RBFN), MLP, SVM and Rotation Forest (RoF). GA with RoF achieved the highest classification accuracy of 99.48%.

Liu et al. [17] proposed predictive models for breast cancer survivability using the DT algorithm on an imbalanced Surveillance, Epidemiology, and End Results (SEER) dataset. In the feature selection stage, LR backward selection was implemented for data reduction and an undersampling method for data balancing. To increase the predictive performance of the classification, the bagging algorithm was implemented and the Area under curve (AUC) results obtained are 76.78%. Miri et al. [18] used the SEER dataset to predict the survivability of patients with breast cancer. To solve the class imbalance problem in the dataset, two oversampling methods such as Borderline-Synthetic Minority Oversampling Technique (SMOTE) and Density-based Synthetic Oversampling (DSO) were used. For feature selection a combination of Correlation-based Feature Selection (CFS) and Particle Swarm Optimisation (PSO) was used. DT, BN and LR were used as classifiers for prediction and an accuracy of 94.33% was achieved when DSO with CFS-PSO was used with DT. Aavula et al. [19] used MLP, DT, LR and SVM classifiers on SEER datasets to predict the survivability of patients with breast cancer. For feature selection, entropy and gain were used to select relevant features using Representative Feature Subset Selection (RFSS) algorithm on the classifiers. SVM-RFSS produced a higher accuracy of 96.78%. Zand et al. [20] predicted breast cancer survivability on the SEER breast cancer dataset by using three classification techniques such as NB, MLP and DT. IG was used to rank features. The DT produced a prediction accuracy of 86.7%. Manikandan et al. [21] used supervised classifiers such as DT, NB and ensemble learning techniques such as AdaBoost, XGBoost and Gradient Boosting classifier for the classification of breast cancer. To select the features Variance Threshold (VT) and PCA was used. The results showed that DT performed better with an accuracy of 98%. Simsek et al. [22] constructed a hybrid DM based methodology to differentiate the importance of variables for survival change over time for three different time periods: 1 year, 5 years and 10 years on the SEER dataset. Least Absolute Shrinkage and Selection Operator (LASSO) technique and GA were used for the independent feature selection. Since the data sets are unbalanced, to balance the number of living and deceased labels in the dataset, two resampling methods such as Random Under Sampling (RUS) and SMOTE were applied. ANN and LR, were applied along with ten-fold cross-validation technique to determine and evaluate the performance of the classification model. A performance analysis was conducted for each model to identify the importance of each variable in the model for time periods of 1 year, 5 years and 10 years and the accuracy obtained has a maximum value of 84%.

Based on the above literature review, it was observed that in the majority of the papers, various feature selection approaches have been explored to improve the accuracy of predicting breast cancer or the survivability of breast cancer patients using the WDBC, WBCD and SEER datasets. This paper proposes a new filter and wrapper sequence using filter methods such as MI, CS and wrapper methods: Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) for the clinical breast cancer dataset, SEER for the classification of the survivability of breast cancer patients as living and deceased with improved accuracy. A breast cancer patient at a later stage of the disease has to pass through mental and physical trauma when subjected to heavy dose chemotherapy or radiotherapy [23]. Upon determining the criticality of breast cancer for survivability, patients can be relieved from such trauma by deciding between second-line treatment or ending the treatment. A second-line palliative or hospice treatment can improve quality of life [24]. So, optimal independent features are identified from the SEER dataset in this paper for predicting the survivability of breast cancer patients as the intersection of independent features obtained from filter methods, MI, CS which are then applied as input to the wrapper method, SFS or SBS. The informative features obtained as output from the wrapper method are subjected to ML algorithms: Multiple Logistic Regression (MLR), K-NN, DT, RF, SVM and MLP. The results are compared using different ML classifiers based on Accuracy (ACC), False Positive Rate (FPR), False Negative Rate (FNR) and Area under the Receiver Operating Characteristics curve (AUC-ROC). The remainder of this paper is organized into four sections. In Section 2, the methodology for the new filter and wrapper sequence is discussed. The experimental results and discussion are presented in Section 3. Finally, Section 4 concludes the paper.

2. PROPOSED METHODOLOGY

The paper proposes a new filter and wrapper sequence to obtain an optimal informative set of independent features for the classification of survivability of breast cancer patients using the clinical breast cancer dataset, SEER and a comparative performance analysis of different ML models in terms of ACC, FPR, FNR and AUC-ROC. The workflow of the proposed filter and wrapper sequence is shown in Fig.1 and discussed in the following subsections.

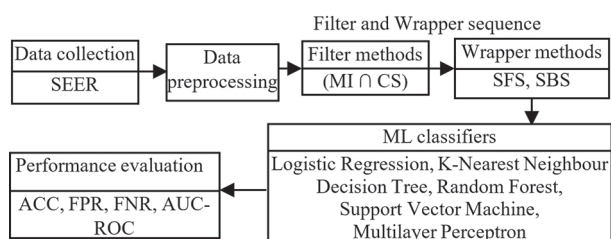


Fig.1. Workflow of the proposed filter and wrapper sequence

2.1. DATA COLLECTION

The SEER dataset-Surveillance, Epidemiology, and End Results is taken from the website, www.seer.cancer.gov. The dataset is an authentic source for cancer statistics updated every year by the Surveillance Research Program (SRP) of Cancer Control and Population Sciences (DCCPS), a division of the National Cancer Institute (NCI), USA. There are 7,755,157 records and 258 features in the dataset collected between the years, 1975 and 2018. The records which have breast cancer information are taken for analysis and so after removing all other types of cancer records only 1,073,477 breast cancer records with 20 features are considered for the later stages of data preprocessing, feature selection, training and testing. The independent features selected from SEER for the classification of survivability of breast cancer patients are age, sex, grade, primary tumour laterality, summary stage, surgery, tumour size, nodes examined, nodes positive, oestrogen receptor (ER) status, progesterone (PR) status, human epidermal growth factor receptor 2 (HER2) status, and the dependent feature is the survival status with labels, living and deceased.

2.2. DATA PREPROCESSING

The datasets collected are preprocessed through steps such as data transformation, data cleaning and data balancing [25]. In the data transformation, the downloaded SEER dataset is transformed into a .csv format as required for implementation in Python. In the data cleaning step, missing records are removed to improve the information content of the data. After cleaning the data, 10,838 data records with 12 features are obtained. The dataset has both categorical and numerical features. The nominal categorical feature values are converted into numerical values by a one-hot encoding technique [26] to work in the Python scikit-learn library. After applying one-hot encoding, the SEER dataset produces 10,838 records with 22 features. The SEER dataset is highly imbalanced. The dataset must be balanced before it is applied to classifiers. In the data balancing stage, the SMOTE-Edited Nearest Neighbour (SMOTE-ENN) [27] technique is applied to handle the imbalance in the training set. SMOTE-ENN is a combination of SMOTE and ENN where SMOTE is an oversampling technique that generates synthetic data of minority samples according to their nearest neighbours. ENN performs data cleaning. After data balancing, 8,769 records corresponding to the majority class, living and 10,233 records corresponding to the minority class, deceased are obtained for the SEER dataset. These records are then subjected to the new filter and wrapper sequence.

2.3. FILTER AND WRAPPER SEQUENCE

The data records obtained after data balancing are subjected to the new filter and wrapper sequence where the optimal independent features are identified with the filter methods: MI, CS and the wrapper meth-

ods: SFS or SBS for the classification of the survivability of breast cancer patients as living and deceased. The intersection of independent features obtained from MI and CS is then applied individually to the SFS and SBS wrapper methods. The filter and wrapper methods are briefed in the following subsections.

2.3.1 Filter Methods

The filter methods used in the new filter and wrapper sequence are Mutual Information and Chi-Square.

Mutual Information

Mutual Information is a filtering method which determines the correlation between independent and dependent features. If each of the n independent features is defined as S_i and the dependent feature as T , then the formula to calculate the Mutual Information, $MI(S_i, T)$ between S_i and T is defined in equation 1.

$$MI(S_i, T) = H(S_i) + H(T) - H(S_i|T) \quad (1)$$

where $1 \leq i \leq n$, $H(S_i)$ is the entropy of independent feature S_i , $H(T)$ is the entropy of dependent class T , and $H(S_i|T)$ is the conditional entropy of S_i and T . The higher values of MI specifies that the independent feature, S_i contains more information for classification [28]. Therefore, k_1 number of independent features whose MI values are greater than 0 are selected to take part in the successive stages of the filter wrapper sequence. The MI values of the independent features obtained by applying the MI filter in this work are listed in Table 1.

Table 1. Independent features selected after applying the MI filter

S.No	Independent features	MI values
1.	Summary stage-localized	0.1046
2.	Age	0.1004
3.	Tumour size	0.0924
4.	ER status-positive	0.0891
5.	PR status-positive	0.0773
6.	Grade-2	0.0722
7.	Laterality-right	0.0634
8.	Nodes positive	0.0554
9.	Grade-1	0.0548
10.	Nodes examined	0.0363
11.	HER2 status-positive	0.0250
12.	HER2 status-negative	0.0165
13.	Laterality-left	0.0135
14.	Surgery-surgery performed	0.0104
15.	Grade-3	0.0096
16.	Sex-female	0.0089
17.	Summary stage-regional	0.0076
18.	ER status-negative	0.0034
19.	Summary stage-distant	0.0008

Chi-Square

Chi-Square is a statistical method which evaluates the independence of the features in a dataset. In this technique, the independence of two events namely the occurrence of independent features and the occurrence of dependent features are evaluated [29]. The equation to calculate the chi-square value, χ^2 for each independent feature is defined in equation 2.

$$CS = \sum_{f=1}^p \frac{(O_f - E_f)^2}{E_f} \quad (2)$$

where O_f is the frequency of the different possible p combinations of S_i independent feature values and T dependent feature values and $1 \leq f \leq p$. E_f is the expected frequency of association between the f^{th} combination of independent feature values and dependent feature values as defined in equation 3. Here Tr_{S_i} is the sum of the records corresponding to each value of the i th independent feature, S_i under consideration, Tc_T is the sum of the records corresponding to each value of the dependent feature, T and m is the total number of records in the training set. When the CS score is higher than the chi-square value, χ^2 , determined from the chi-square distribution table, corresponding to the degrees of freedom, dof, the features are highly related. B_{S_i} in the independent feature, S_i and the number of values in the dependent feature B_T . The dof is calculated as the product of B_{S_i} and B_T as in equation 4.

$$E_f = \frac{Tr_{S_i} \times Tc_T}{m} \quad (3)$$

$$dof = (B_{S_i} - 1) \times (B_T - 1) \quad (4)$$

If the CS score is lower than the χ^2 score, the features are less correlated. The independent features with low CS scores are not included when modelling the classifier. The CS values of the independent features obtained after applying the CS filter in the proposed work are listed in Table 2.

Table 2. Independent features selected after applying the CS filter

S.No	Independent features	CS values
1.	Tumour size	28640.032
2.	Nodes positive	8185.748
3.	Age	5188.570
4.	Nodes examined	3048.862
5.	Grade-2	1407.698
6.	Summary stage-localized	1329.466
7.	Grade-1	1255.024
8.	Laterality-right	1194.877
9.	PR status-positive	1030.023
10.	ER status-positive	717.781
11.	HER2 status-positive	638.850
12.	Laterality-left	231.628
13.	Summary stage-regional	204.313

S.No	Independent features	CS values
14.	HER2 status-negative	82.197
15.	Grade-3	70.376
16.	Sex-male	39.709
17.	ER status-negative	25.409
18.	PR status-negative	17.142
19.	Summary stage-distant	4.046
20.	Sex-female	2.078
21.	Surgery-not performed	1.384
22.	Surgery-surgery performed	0.922

The intersection of independent features obtained from both filter methods, *MI* and *CS* is used to find the optimal set of features. The independent features, Summary stage-localized, ER status-positive, Nodes examined, Laterality-right, HER2 status-negative, HER2 status-positive, Tumour size, Surgery-surgery performed, Grade-1, Laterality-left, Age, Summary stage-regional, ER status-negative, PR status-positive, Grade-2, Summary stage-distant, Sex-female, Grade-3 and Nodes positive obtained are subjected as inputs to the wrapper methods: SFS or SBS to find a more optimal set of independent features from the SEER training set for the classification of the survivability of breast cancer patients.

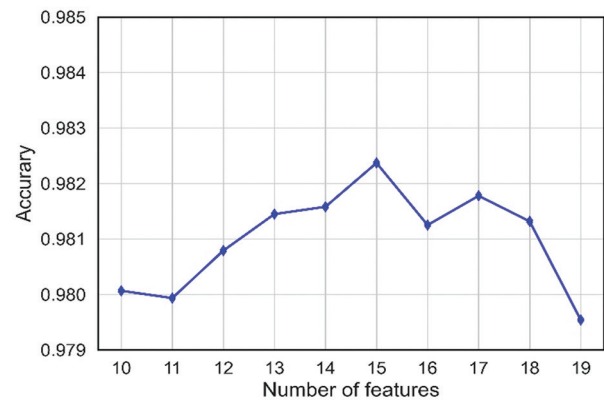
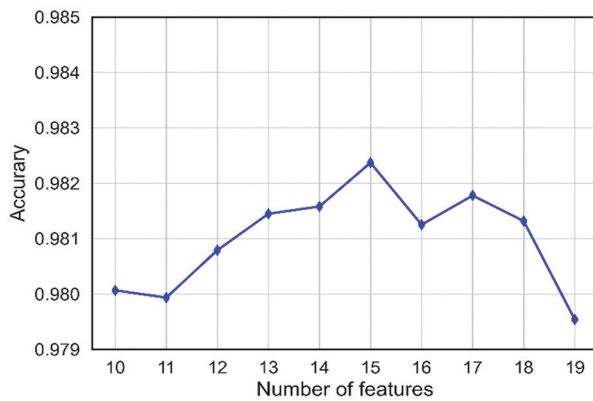


Fig. 2. Line plot between independent features and the accuracy scores a) SFS b) SBS

Table 3. Independent Features obtained after applying wrapper methods, SFS and SBS

SFS	SBS
Summary stage-localized	Summary stage-localized
Nodes examined	ER status-positive
Laterality-right	Nodes examined
HER2 status-negative	Laterality-right
HER2 status-positive	Tumour size
Tumour size	Grade-1
Surgery-surgery performed	Laterality-left
Grade-1	Age
Laterality-left	Summary stage-regional
Age	ER status-negative
Summary stage-regional	Grade-2
ER status-negative	Summary stage-distant
PR status-positive	Sex-female
Summary stage-distant	Grade-3
Sex-female	Nodes positive

The main objective of the new filter and wrapper sequence is to remove irrelevant independent features from the training set and reduce the dimension of the training set. From the independent feature subset obtained from the wrapper methods, SFS and SBS, the first 10 and 15 independent features are selected and subjected to ML algorithms to analyse the performance of classifiers. The independent features that are distinctly identified by SFS are HER2 status-positive, HER2 status-negative, Surgery-surgery performed and PR status-positive and by SBS are ER status-positive, Grade-2, Grade-3 and Nodes positive.

2.4. CLASSIFIERS

Classification is a supervised learning algorithm that identifies the value of dependent feature of a given independent test data record based on the classifier

model produced from a labelled training set. In binary classification, the classifier obtained after the training phase predicts one of the two values of the dependent feature. Six ML based binary classifiers such as MLR, K-NN, DT, RF, SVM and MLP are used to analyse the performance of the new filter and wrapper sequence for predicting the survivability of breast cancer patients as living and deceased. Multiple Logistic Regression (MLR) [31] predicts the probability of a dependent feature using a logistic function. To identify the value of a dependent feature, the threshold value set corresponding to the different independent features is determined during the training phase of the algorithm. The test record whose values lie above the threshold value set falls into one class and the test record whose values which lie below falls into another class. K-Nearest Neighbour (K-NN) [32] identifies the class of a test record by using the dependent feature values of 'K' neighbours nearest to the test record under consideration. The K-nearest neighbours are identified using the distance measures such as Euclidean, Manhattan, Minkowski or Hamming distance. Decision Tree (DT) [33] determines the relationship between independent and dependent features in the form of a tree like structure based on measurements of information content in the independent features. The branches of the DT represent a decision rule set for identifying the class of the test record. Random Forest (RF) [34] is a top-down approach in which a number of decision trees are obtained using various subsets of the training set and the ensemble of their results is predicted as the dependent feature value of the incoming test records. The more the number of decision trees in the forest, the greater the accuracy. Support Vector Machine (SVM) [35] segregates the n-dimensional space of the independent features of the training set by an optimal hyperplane which lays the maximum distance between the support vectors on either side. The parameters of this optimal hyperplane help classify the incoming test records. Multilayer Perceptron (MLP) [36] is a feed forward neural network where the weights of the links connecting the input and the hidden layer, hidden and output layer are optimised during back propagation-based training to obtain an optimised weight vector which can predict the dependent feature value of the test record. The performance of these ML classifiers is evaluated using the objective metrics namely ACC, FPR, FNR and AUC-ROC. The results are compared and discussed in the next section.

3. EXPERIMENTAL RESULTS AND DISCUSSION

In the proposed new filter and wrapper sequence, the SEER clinical breast cancer dataset is used in the experimental analysis. The training set is subjected to the data preprocessing stage and the new filter and wrapper sequence. In the data preprocessing step, data transformation, data cleaning and data balancing steps are carried out. The filter methods, MI and CS are applied to the preprocessed training sets. The intersec-

tion of more relevant independent features identified from the filter stage are subjected to the SFS and SBS wrapper methods to obtain a more optimal set of independent features. The optimal set of 10 and 15 independent features are applied to the ML algorithms and the performance of the analysis are evaluated in this section in terms of the evaluation metrics: ACC, FPR, FNR and AUC-ROC. These evaluation metrics are defined in the following subsection.

3.1. EVALUATION METRICS

The performance evaluation of the classifiers is mainly based on the correct and incorrect predictions made by the model. The confusion matrix provides more insight into the performance of a prediction model and also identifies the classes which are correctly and incorrectly predicted by the model. Accuracy provides the ratio of the number of correct predictions to the total number of predictions made by the model. False Positive Rate refers to the number of predictions where the classifier incorrectly predicted the deceased class as living [18]. False Negative Rate refers to the number of predictions where the classifier incorrectly predicted the living class as deceased [15]. AUC-ROC curve is drawn by plotting the FPR and TPR values. The curve is plotted for different probability thresholds of the models while predicting the probability of different classes [18]. The ROC curve corresponding to the largest area has a better ability to classify between living and deceased classes.

3.2. RESULTS

The values of different evaluation metrics obtained from the testing stage are tabulated in Table 4. The results obtained show that, when the number of independent features identified from the new filter wrapper sequence is 15, SVM produced an accuracy of 99% and DT, an accuracy of 98.1% from SFS. Similarly, when the selected independent features are 10, DT has an accuracy of 85.9%. In the case of the SBS wrapper technique, SVM yields the highest accuracy of 99.5% and K-NN produced an accuracy of 98.7% when 15 selected features are used. When 10 independent features are selected, DT obtained an accuracy of 86.3%. SBS performed better than SFS across all ML algorithms when 15 independent features are selected. When SBS was used, K-NN produced an FPR of zero, and when SFS was used with 15 independent features, it produced an FPR of 0.001. When 15 independent features were used, SVM produced an FNR value of 0.005 with SBS and 0.006 with SFS. Based on the results, K-NN and SVM perform better in terms of FPR and FNR. The accuracy values obtained from different classifiers for 10 and 15 selected independent features from the wrapper technique, SFS and SBS are shown in the Fig. 3 and Fig. 4. Comparatively, the SBS wrapper technique produced higher accuracy than the SFS.

Table 4. Comparative analysis on different evaluation metrics using SEER dataset

Classifiers	Features=10							
	SFS				SBS			
	ACC	FPR	FNR	AUC-ROC	ACC	FPR	FNR	AUC-ROC
MLR	83.4	0.290	0.032	89.9	83.7	0.168	0.158	90.9
K-NN	84.5	0.151	0.159	92.3	84.2	0.132	0.186	92.7
DT	85.9	0.220	0.055	93.4	86.3	0.143	0.131	93.4
RF	85.8	0.219	0.059	93.4	85.7	0.133	0.153	93.7
SVM	85.7	0.234	0.043	90.9	86.1	0.127	0.153	92.4
MLP	85.7	0.223	0.057	93.1	85.3	0.144	0.150	93.4
Features=15								
MLR	89.3	0.156	0.053	95.8	90.7	0.110	0.074	97.4
K-NN	97.9	0.001	0.044	99.3	98.7	0	0.027	99.5
DT	98.1	0.013	0.026	98.1	97.9	0.012	0.030	97.9
RF	96	0.035	0.045	99.4	96.9	0.028	0.033	99.6
SVM	99	0.006	0.014	100	99.5	0.005	0.005	100
MLP	90.3	0.111	0.082	96.6	91.5	0.090	0.079	97.6

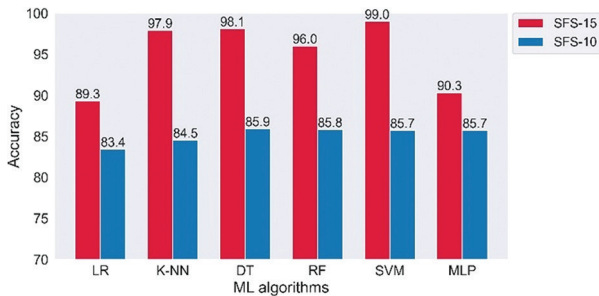


Fig. 3. Accuracy values of different ML algorithms using SFS

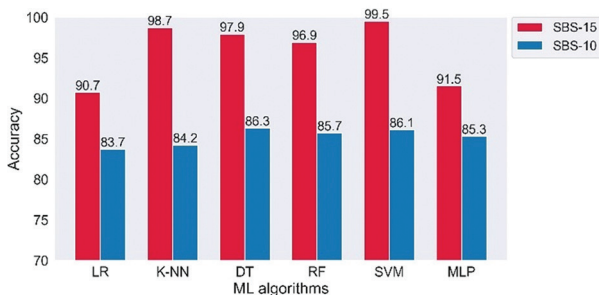


Fig. 4. Accuracy values of different ML algorithms using SBS

The ROC curves are drawn between the FPR and TPR. When the model predicts the probability of belonging to different classes, curves are plotted for different thresholds of the ML models under comparison. The ROC curves are plotted between FPR and TPR for the classifiers, MLR, K-NN, DT, RF, SVM and MLP which correspond to the proposed feature selection sequence using SFS and SBS for 15 independent features from the SEER dataset, as shown in Fig.5 and Fig.6. The AUC-ROC curves are higher for the ML classifiers when 15 independent features identified from the proposed feature

selection sequence are used. According to Fig. 5 and Fig. 6, the AUC-ROC curve of the SVM classifier is larger for both SFS and SBS.

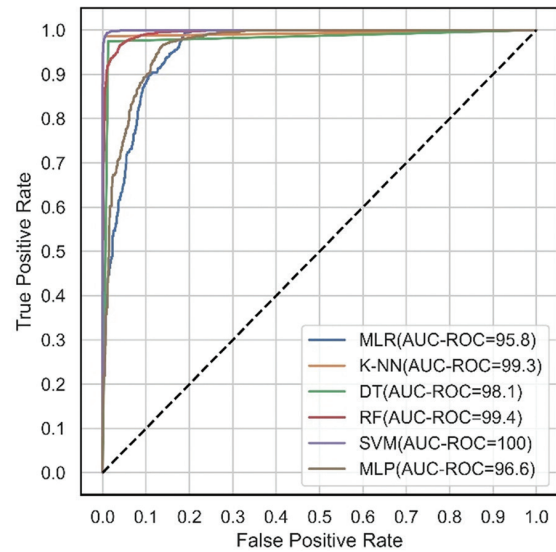


Fig. 5. ROC curves for the proposed feature selection sequence-SFS with 15 independent features

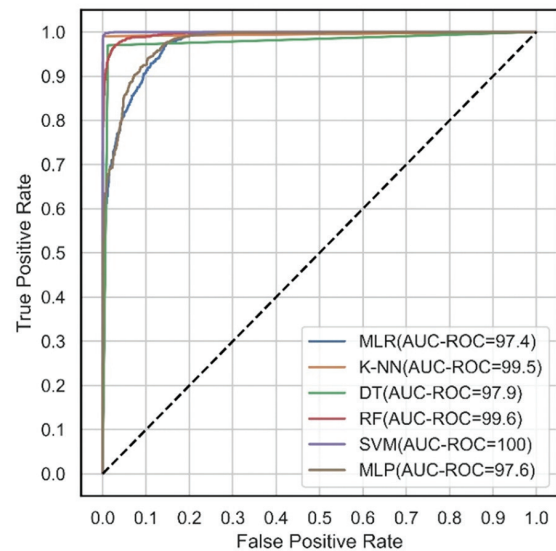


Fig. 6. ROC curves for the proposed feature selection sequence-SBS with 15 independent features

Table 5. Comparison with results from other features selection techniques-SEER dataset

References	Number of features used	Feature selection techniques	ML which produces highest ACC & AUC (%)
Liu et al. [17]	11	LR backward selection	DT-76.78 (AUC)
Miri et al. [18]	10	CFS, PSO	DT-94.33
Zand et al. [20]	16	IG	DT-86.7
Manikandan et al. [21]	13	VT, PCA	DT-98
Proposed Method	10 and 15	Filter: MI, CS, Wrapper: SBS	SVM-99.5

In addition, the results produced by the proposed feature selection sequence is compared with the results obtained in previous studies [17-22] which use the respective feature selection sequences as mentioned in Table 5. The accuracy produced using the feature selection techniques in Table 5 based on SEER are less than the results produced by the proposed feature selection sequence.

4. CONCLUSION

In this paper, a new filter and wrapper feature selection sequence based on the filter methods, MI, CS and SFS, SBS is proposed. The intersection of independent features obtained from both the filter methods, MI and CS is used to find the optimal set of features. The independent features obtained are subjected as input to the wrapper methods, SFS or SBS to determine a more optimal set of independent features from the SEER training set for the classification of the survivability of breast cancer patients. The results show that SVM performed better than other algorithms, with 99.5% accuracy and higher AUC-ROC values. When SBS and 15 independent features were used, K-NN and SVM both produced lower FPR and FNR values. Compared to SFS, SBS produced better results when 15 independent features are selected. In addition, the results are compared with those obtained using other feature selection techniques in the SEER dataset. It is found that the proposed feature selection sequence with SBS produced higher values for all evaluation metrics when compared to other feature selection techniques in the comparative study while predicting the survivability of breast cancer patients.

5. REFERENCES:

- [1] R. D. Kehm, A. A. Llanos, J. A. McDonald, P. Tehranifar, M. B. Terry, "Evidence-Based Interventions for Reducing Breast Cancer Disparities: What Works and Where the Gaps Are?", *Cancers*, Vol. 14, No. 17, 2022, p. 4122.
- [2] C.A.U. Hassan, M.S. Khan, M.A. Shah, "Comparison of machine learning algorithms in data classification", *Proceedings of the 24th International Conference on Automation and Computing*, Newcastle Upon Tyne, UK, 6 September 2018, pp. 1-6.
- [3] P. Misra, A. S. Yadav, "Impact of preprocessing methods on healthcare predictions", *Proceedings of the 2nd International Conference on Advanced Computing and Software Engineering*, Sultanpur, India, 9 March 2019, pp. 144-150.
- [4] N. Somu, M. G. Raman, K. Kirthivasan, V. S. Sriram, "Hypergraph based feature selection technique for medical diagnosis", *Journal of Medical Systems*, Vol. 40, 2016, pp. 1-16.
- [5] K. N. Neeraj, V. Maurya, "A review on machine learning (feature selection, classification and clustering) approaches of big data mining in different area of research", *Journal of Critical Reviews*, Vol. 7, No. 19, 2020, pp. 2610-2626.
- [6] V. Kumar, G. S. Lalotra, P. Sasikala, D. S. Rajput, R. Kaluri, K. Lakshmana, M. Shorfuzzaman, A. Alsufyani, M. Uddin, "Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques", *Healthcare*, Vol. 10, No. 7, 2022, p. 1293.
- [7] E. O. Omuya, G. O. Okeyo, M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain", *Expert Systems with Applications*, Vol. 174, 2021, p. 114765.
- [8] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction", *Journal of Applied Science and Technology Trends*, Vol. 1, No. 2, 2020, pp. 56-70.
- [9] D. H. Jeong, B. K. Jeong, N. Leslie, C. Kamhoua, S. Y. Ji, "Designing a supervised feature selection technique for mixed attribute data analysis", *Machine Learning with Applications*, Vol. 10, 2022, p. 100431.
- [10] Y. M. Sobhanzadeh, H. Motieghader, A. Masoudi-Nejad, "Feature Select: a software for feature selection based on machine learning approaches", *BMC Bioinformatics*, Vol. 20, No. 1, 2019, pp. 1-7.
- [11] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevinoo, J. Tang, H. Liu, "Feature selection: A data perspective", *ACM Computing Surveys*, Vol. 50, No. 6, 2017, pp. 1-45.
- [12] D. M. Liou, W. P. Chang, "Applying Data Mining for the Analysis of Breast Cancer Data", *Data Mining in Clinical Medicine*, Vol. 1246, 2015, pp. 175-189.
- [13] A. Saygili, "Classification and Diagnostic Prediction of Breast Cancers via Different Classifiers", *International Scientific and Vocational Studies Journal*, Vol. 2, No. 2, 2018, pp. 48-56.
- [14] D. A. Omondiagbe, S. Veeramani, A. S. Sidhu, "Machine Learning Classification Techniques for Breast Cancer Diagnosis", *IOP Conference Series: Materials Science and Engineering*, Vol. 495, No. 1, 2019, p. 012033.
- [15] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, M. N. Kabir, "Breast cancer prediction: a comparative study using machine learning techniques", *SN Computer Science*, Vol. 1, No. 5, 2020, pp. 1-14.

- [16] E. Alickovic, A. Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest", *Neural Computing and Applications*, Vol. 28, No. 4, 2015, pp. 753-763.
- [17] Y. Q. Liu, C. Wang, L. Zhang, "Decision Tree Based Predictive Models for Breast Cancer Survivability on Imbalanced Data", *Proceedings of the 3rd International Conference on Bioinformatics and Biomedical Engineering*, Beijing, China, 11-13 June 2009, pp. 1-4.
- [18] S. M. Rostami, M. Ahmadzadeh, "Extracting predictor variables to construct breast cancer survivability model with class imbalance problem", *Journal of AI and Data Mining*, Vol. 6, No. 2, 2018, pp. 263-276.
- [19] R. Aavula, R. Bhramaramba, "XBPF: an extensible breast cancer prognosis framework for predicting susceptibility, recurrence and survivability", *International Journal of Engineering and Advanced Technology*, Vol. 8, No. 5, 2019, pp. 2249-8958.
- [20] H. Karim, K. Zand, "A comparative survey on data mining techniques for breast cancer diagnosis and prediction", *Indian Journal of Fundamental and Applied Life Sciences*, Vol. 5, No. 1, 2015, p. 4330.
- [21] P. Manikandan, U. Durga, C. Ponnuraja, "An Integrative Machine Learning Framework for Classifying SEER Breast Cancer", <https://ssrn.com/abstract=4308284> (accessed: 2022)
- [22] S. Simsek, U. Kursuncu, E. Kibis, M. A. Abdellatif, A. Dag, "A hybrid data mining approach for identifying the temporal effects of variables associated with breast cancer survival", *Expert Systems with Applications*, Vol. 139, 2019, p. 112863.
- [23] A. Lahousse, E. Roose, L. Leysen, S. T. Yilmaz, K. Mostaqim, F. Reis, J. Nijs, "Lifestyle and pain following cancer: State-of-the-art and future directions", *Journal of Clinical Medicine*, Vol. 11, No. 1, 2021, p. 195.
- [24] M. Petrova, G. Wong, I. Kuhn, I. Wellwood, S. Barclay, "Timely community palliative and end-of-life care: a realist synthesis", *BMJ Supportive and Palliative Care*, 2021, pp. 1-15.
- [25] A. Idri, H. Benhar, J. L. Fernandez-Aleman, I. Kadi, "A systematic map of medical data preprocessing in knowledge discovery", *Computer Methods and Programs in Biomedicine*, Vol. 162, 2018, pp. 69-85.
- [26] R. Gupta, R. Bhargava, M. Jayabalan, "Diagnosis of Breast Cancer on Imbalanced Dataset Using Various Sampling Techniques and Machine Learning Models", *Proceedings of the 14th International Conference on Developments in eSystems Engineering*, Sharjah, United Arab Emirates, 7-10 December 2021, pp. 162-167.
- [27] S. Fotouhi, S. Asadi, M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data", *Journal of Biomedical Informatics*, Vol. 90, 2019, p.103089.
- [28] F. Souza, C. Premebida, R. Araujo, "High-order conditional mutual information maximization for dealing with high-order dependencies in feature selection", *Pattern Recognition*, Vol. 131, 2022, p. 108895.
- [29] K. Juneja, C. Rana, "An improved weighted decision tree approach for breast cancer prediction", *International Journal of Information Technology*, Vol. 12, No. 3, 2020, pp. 797-804.
- [30] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, S. Fong, "Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy", *Pertanika Journal of Science and Technology*, Vol. 26, No. 1, 2018, pp. 329-340.
- [31] C. Shravya, K. Pravalika, S. Subhani, "Prediction of breast cancer using supervised machine learning techniques", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 8, No. 6, 2019, pp. 1106-1110.
- [32] R. Ehsani, F. Drablos, "Robust Distance Measures for k NN Classification of Cancer Data", *Cancer Informatics*, Vol.19, 2020, pp.1-9.
- [33] R. Ahuja, A. Chug, S. Gupta, P. Ahuja, S. Kohli, "Classification and clustering algorithms of machine learning with their applications", *Nature-inspired Computation in Data Mining and Machine Learning*, Vol. 855, 2020, pp. 225-248.
- [34] M. Schonlau, R. Y. Zou, "The random forest algorithm for statistical learning", *The Stata Journal*, Vol.20, No.1, 2020, pp. 3-29.
- [35] S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. K. Thong, J. S. Polster, B. Steiert, J. D. Zhang, "An introduction to machine learning", *Clinical Pharmacology and Therapeutics*, Vol. 107, No. 4, 2020, pp. 871-885.
- [36] M. Desai, M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)", *Clinical eHealth*, Vol. 4, 2021, pp. 1-11.

A New Approach using Deep Learning and Reinforcement Learning in HealthCare: Skin Cancer Classification

Original Scientific Paper

Dahdouh Yousra

Faculty of Sciences and Techniques, LIST Laboratory FSTT UAE
Tangier, Morocco
dahdouhyousra@gmail.com

Anouar Boudhir Abdelhakim

Faculty of Sciences and Techniques, LIST Laboratory FSTT UAE
Tangier, Morocco
aboudhir@uae.ac.ma

Ben Ahmed Mohamed

Faculty of Sciences and Techniques, LIST Laboratory FSTT UAE
Tangier, Morocco
m.benahmed@gmail.com

Abstract – Nowadays, skin cancer is one of the most important problems faced by the world, due especially to the rapid development of skin cells and excessive exposure to UV rays. Therefore, early detection at an early stage employing advanced automated systems based on AI algorithms plays a major job in order to effectively identifying and detecting the disease, reducing patient health and financial burdens, and stopping its spread in the skin. In this context, several early skin cancer detection approaches and models have been presented throughout the last few decades to improve the rate of skin cancer detection using dermoscopic images. This work proposed a model that can help dermatologists to know and detect skin cancer in just a few seconds. This model combined the merits of two major artificial intelligence algorithms: Deep Learning and Reinforcement Learning following the great success we achieved in the classification and recognition of images and especially in the medical sector. This research included four main steps. Firstly, the pre-processing techniques were applied to improve the accuracy, quality, and consistency of a dataset. The input dermoscopic images were obtained from the HAM10000 database. Then, the watershed algorithm was used for the segmentation process performed to extract the affected area. After that, the deep convolutional neural network (CNN) was utilized to classify the skin cancer into seven types: actinic keratosis, basal cell carcinoma, benign keratosis, dermatofibroma melanocytic nevi, melanoma vascular skin lesions. Finally, in regards to the reinforcement learning part, the Deep Q_Learning algorithm was utilized to train and retrain our model until we found the best result. The accuracy metric was utilized to evaluate the efficacy and performance of the proposed method, which achieved a high accuracy of 80%. Furthermore, the experimental results demonstrate how reinforcement learning can be effectively combined with deep learning for skin cancer classification tasks.

Keywords: Skin Cancer, Deep Learning, Reinforcement Learning, Classification, CNN, Deep Q_Learning, Dermoscopy Image, Segmentation

1. INTRODUCTION

Cancer is a disease caused by the transformation of cells that become abnormal and proliferate excessively, it is one of the main causes of death worldwide, with approximately 10 million deaths in 2020 [1]. The majority of human cancers are skin cancers, and their prevalence is increasing more quickly than all other cancers [2].

Skin cancer is one of the most serious forms of cancer [3] it arises when skin cells grow irregularly; factors contributing to its occurrence include UV radiation exposure, a family history of the disease, decreased immunity, etc. Morocco reported 114 fatalities and 248 new cases of melanoma in 2020 [4].

The three main types of skin cancer are: Basal cell carcinoma, Squamous cell carcinoma, and Melanoma,

this last is regarded as the most hazardous form of all the other varieties. Early and accurate detection is considered the best method of surviving and avoiding the worst effects of cancer and cured. But that is a difficult task since malignant tumors and normal moles share visual characteristics, which makes it difficult, particularly for the detection of type melanoma [5].

Dermoscopic imaging methods are most frequently employed by dermatologists to examine skin lesions based on their collection of morphological characteristics. But only qualified medical professionals with the proper vision and experience can employ this approach to its full potential [6]. These difficulties have recently presented the research community with a major challenge to develop new and innovative systems based on artificial intelligence (AI) tools including those based on computer algorithms, deep learning, and deep reinforcement learning networks [7] for aided assist experts to early diagnosing of skin cancer and prevention.

Artificial intelligence (AI) is a branch of computing dedicated to the design of machines capable of imitating human intelligence that relies on the creation and application of algorithms executed in a dynamic computing environment. Its purpose is to allow computers to think and act like human beings. With advances the last few years have seen in computing and information science, artificial intelligence (AI) is rapidly becoming an integral part of modern healthcare. AI algorithms are used to help healthcare professionals in clinical settings, in ongoing research, and in diagnostic dermatology [8]. The most popular algorithms used are deep learning, machine learning, and reinforcement learning.

Deep learning is a branch of machine learning, which is based on a group of algorithms inspired by the human brain (ANN) that seek to shape high-level abstractions of data using multiple layers (input layer, hidden layer, output layer). It has been applied in several areas such as speech recognition [9], bioinformatics [10], and also detection/diagnosis in medical imaging [11]. Recently, the convolutional neural network (CNN) is considered the most successful algorithm applied for major medical image tasks, such as image classification, segmentation, localization, and detection.

Reinforcement Learning is the most popular method of machine learning that handles sequential decision problems [12], it involves letting computers learn from their experiences through a reward or penalty system. Reinforcement learning has achieved very success in recent years in the following artificial intelligence applications: robot control, computer vision, autonomous driving, and computer gaming. And also, has emerged as one of the crucial areas in the field of artificial intelligence impacting the field of health care including diagnosis, prognosis, and other medical treatments [13].

Deep Q_Networks (DQNs) [14] are neural networks that use deep Q_learning to provide models, it's been

composed of convolutional neural networks and other structures that use specific methods to learn more about various processes with high accuracy. DQN can be effectively used to detect skin cancer [15].

In this work, we have proposed a new and intelligent model to increase the diagnostic accuracy of skin cancer. This model combines the merits of two major artificial intelligence algorithms: Deep Learning (CNN) and Reinforcement Learning (DQN). We have collected 10015 dermoscopy images from the ISIC site archive, which contains 7 types of skin cancers. firstly, the pre-processing techniques have been applied, such as reading, resizing images, cleaning images, and applying One Hot Encoding on the labels of the dataset and splitting into a training set and a test set. After that, the segmentation process was used to extract the affected area using the watershed algorithm. Then, extracting important features from images and classified the dermoscopy images using the deep convolutional neural network (CNN). And about the reinforcement learning part, we used the Deep Q_Learning algorithm to train and retrain our classification model until we found the best result.

The main contributions of the proposed work are:

- Use of reinforcement learning to classification tasks, and especially to classify dermoscopy images for the first time in the literature.
- This paper proposes a method for detecting skin cancer kinds in dermoscopy images utilizing the ISIC database.
- The proposed model performs significantly better in the classification task.
- Best performance metrics obtained using a deep algorithm contain many layers with different parameters.
- Combined reinforcement learning with deep learning to detect and classification of skin cancer.

The structure of this paper is as follows: Section 2 presents an overview of related literature. Section 3, called materials and methods, first provides a description of the dataset used. Next, presents the theory of deep learning (CNN) and reinforcement learning (DQL), and finally, explains in detail the method proposed. Section 4 contains the experiment results, including the experiment settings and evaluation of the proposed model. Finally, Section 5 concludes the current study, with a little discussion and suggests possible future work avenues.

2. LITERATURE REVIEW

Skin cancer is one of the greatest challenges in the medical domain. Recently, various types of diagnostic techniques based on artificial intelligence tools that can early detect and classify skin cancer have been proposed and developed.

In related works, many researchers have used IA algorithms to automatically diagnose skin diseases.

- **Machine Learning based Approaches:**

In [16] the authors proposed an approach to detect and classify eight types of skin cancer using dermoscopic images collected by the ISIC 2019 challenge dataset. Which has been used principally in the pre-processing phase with Gaussian and median filters for noise removal and image enhancement. Also, the Dull Razor method to remove any extra hair from the skin lesion. After that utilized the clustering algorithm: Color-based k- means clustering to segmentation. To facilitate the classification purposes two methods of extracting features have been applied: ABCD and GLCM. In the end, the authors decide to use part of the support vector machine (MSVM) due to the great success it has achieved to solve multi-class problems. This approach attained good accuracy.

In [17] the authors concentrate on the identification diagnosis of skin cancer melanoma with Supervised Machine Learning using Cubic Regression. This method required various steps like the image processing technique to select of suitable color model for the image being processed, segmentation by creating masking to segment the image, feature extraction, and machine learning with two modes: train and test, and graphical presentation of data have used cubic regression as a good solution for displaying the result. The main objective of this project is to train a machine to automatically display the stages of skin cancer.

In [18] the authors implemented a model based on the KNN algorithm to detect and classify skin lesions (normal or benign). The proposed model consists of four steps: preprocessing involved with two-phase: image enhancement and hair removal, segmentation using the thresholding method, feature extraction, and classification with KNN which is based on calculating the distance between a number of data points in an image. The proposed system showed satisfactory accuracy of classification. In this way, this model was able specifically to classify skin lesions.

- **Transfer Learning-based Approaches:**

In [19] the authors used a deep convolutional neural network (DCNN) to classify skin lesions between benign and malignant based on the HAM10000 dataset that incorporates the following steps: preprocessing with applied some filter to remove noise, data reduction, normalization of data, feature extraction and transformed data label to numeric, data augmentation techniques to the expansion of the data and increase the number of images. the goal of the authors is to compare the performance between DCNN and transfer learning models as VGG16 and other models. The proposed DCNN model achieved a robust result.

In [20] the authors presented a pre-trained Xception model with a method fine-tuned to classify skin cancer by adding a series of layers after the base layer of the Xception model and retraining the model weights. The input images are resized to 224 x 224 pixels. The used loss func-

tion is categorical cross entropy, while the optimizer is Adam with a learning rate of 0.0001. The results obtained indicate that the proposed model is both efficient.

In [21] the authors developed a model that combines pre-trained convolutional neural networks (DenseNet201) architectures as feature extractors and machine learning (Cubic SVM) as a classifier. The choice of these models depends on the results of several training steps on the PH2 dataset. The methodology used contains three principal stages: input images, the feature extraction with compared three pre-trained models: ResNet, DenseNet, and EfficientNet. Classification with compared four algorithms: Artificial Neural Network, Support Vector Machines, K-Nearest Neighbor, and Random Forest. The result obtained shows the great performance of the model to detect skin lesions.

- **Deep Learning based Approaches:**

In [22] the authors proposed a medical artificial solution called the LNet model based on a deep convolutional neural network (DCNN) to classify binary skin cancer using dermoscopy images. This model contained many layers designed using 11 blocks, the block of the network is composed of convolutional, pooling, BN, and leakyReLU layers that make use of a different set of parameters, including the number of kernels, stride, and filter to extract features and classify skin lesions. The authors applied data augmentation to address the problem of lack of data. This method achieved good results in expediting the process of melanoma diagnosis automatically.

In [23] the authors designed an automatic technique based on a deep learning algorithm with Fuzzy K-Means Clustering to detect and identify the kinds of skin cancer. The method proposed in this paper contains three steps: firstly, preprocessing using the morphological closing process and other filters and then applying a faster RCNN to obtain fixed-length feature vectors through four stages: convolution layers, regional proposal networks, and classification. and finally, for segmentation FKM has been used due to it working well for overlapped data. This method has been evaluated by three datasets PH2, ISIC-2017, and ISBI-2016, and the experimental results exhibit the advantage and the performance of using this method in the detection and segmentation of skin lesions.

In [24] to automatically identify skin cancer the authors proposed a system using a convolutional neural network (CNN) containing 3 hidden layers: a convolution layer, a pooling layer, a fully connected layer with softmax activation, and uses multiple optimizers such as Adam, Nadam, SGD, and RMSprop with learning rates of 0.001. The authors found that adam optimizer provided the best performance in identifying skin lesions in 5 categories (i.e., dermatofibroma, nevus pigmentosus, squamous cell carcinoma, and melanoma) from the ISIC dataset after using augmentation techniques. The objective of the study is to obtain the best performance of the skin cancer classification methods existing.

- **Overview of our proposed approach:**

The idea behind our approach is to build an efficient and intelligent system to assist doctors in skin cancer detection, increase diagnostic accuracy, and make patients' lives easier. The system proposed combines two major artificial intelligence algorithms: Deep Learning (CNN) and Reinforcement Learning (DQN) due to the great advantages to learn high-level features from data, and innovation to do the task of classification and detection.

The novelty of this research manifests itself in the following:

- How to effectively combined reinforcement learning with deep learning. And how deep reinforcement learning can help in the medical context, especially in the diagnosis and detection of skin cancer.
- A novel approach to the classification and detection the skin cancer is proposed which is based on deep reinforcement learning. This is a first to the best of my knowledge.
- For the first time used reinforcement learning for the classification task and especially to classify the dermoscopic images.
- The architecture of the two algorithms is deep and contains many layers with different parameters.
- This paper addresses the issue of investigating the performance of the CNN-DQN model for the classification and detection of skin lesions.
- The results explain the success of our proposed approach, as we found satisfactory accuracy.

Our proposed approach contains four steps: (1) The input dermoscopy images passed in the preprocessing process to remove the noise, cleaning, normalization, and splitting data. (2) Extract the affected area with applied the watershed algorithm to the output of the preprocessing phase. (3) classification using the deep convolutional neural network (CNN) model containing a sequence of convolutional layers, pooling layers, several activation functions, Batch Normalization, Dropout, and fully connected layer "Softmax function". (4) Training phase using the reinforcement learning (Deep Q_Learning algorithm) until to obtain the best result.

3. MATERIALS AND METHODS

In this section, we present the dataset used first and then a description of two algorithms used: deep learning "CNNs" and reinforcement learning "DQN", and finally we provide the detail about our proposed model.

3.1. DATA DESCRIPTION

To conduct this research and analyze the experimental findings of the suggested strategy. We have used the publicly available HAM10000 dataset. This dataset contains 10015 dermoscopic images devised of seven different classes of skin lesions as shown in Table 1 [25]. These dermoscopic images were originally 600 x

450 pixels in RGB format, which We then resized all to 28 x 28 pixels to reduce network input and parameters.

We used the Train_Test_Split technique to randomly split the data: 80% for training and 20% for testing.

Table 1. Presents the seven classes and the type of cancer in each class

Class	Type	Number of images
Actinic keratosis	Benign or Malignant	327
basal cell carcinoma	Malignant	541
benign keratosis	Benign	1099
Dermatofibroma	Benign	155
Melanocytic nevi	Benign	6705
Melanoma	Malignant	1113
vascular skin lesions	Benign or Malignant	142

3.2. DEEP LEARNING

Deep learning is a type of artificial intelligence derived from machine learning where the machine is able to learn by itself. Deep Learning is based on a network of artificial neurons inspired by the human brain. Recently, deep learning has been shown to be a particularly effective technology due to its ability to manage massive volumes of data. which it's used for different work such as object detection, image classification, and speech recognition... The convolutional neural network algorithm is the most popular and used, particularly for classifying medical images.

3.3. CNN

CNN is a multi-layer neural network with a special architecture developed to extract increasingly complex data characteristics from each layer in order to determine the output as shown in Fig. 1. Used primarily when an unstructured data set exists [26, 27]. One of the most popular uses of this architecture is image processing, classification, and segmentation.

The layers of a CNN consist of an input layer, an output layer, and a hidden layer that consists of multiple convolutional layers, clustering layers, fully connected layers, and normalization layers.

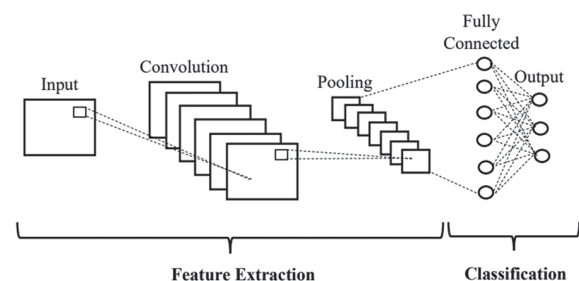


Fig. 1. The convolutional neural network architecture

This work has incorporated the CNN algorithm into diagnostic skin cancer with dermoscopic images, where we implemented a deep CNN architecture for the classification task.

3.3. REINFORCEMENT LEARNING

An agent and an environment are the two principal elements of the machine learning technique known as reinforcement learning (RL) [28]. This type of learning is based on interaction with the environment by trial and error using feedback from its own actions and experiences. The main objective of reinforcement learning is to find the most appropriate action model for maximizing the total cumulative rewards of RL agents.

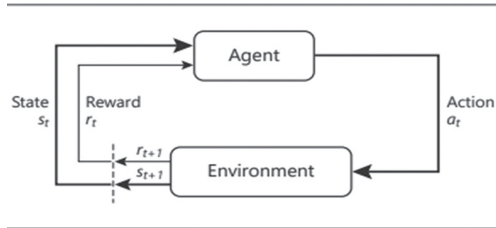


Fig. 2. Illustration of the agent-environment interface [28].

In recent years, reinforcement learning has emerged as one of the crucial areas in the field of artificial intelligence impacting the field of health care. Deep Reinforcement Learning (DRL) combines deep neural networks and reinforcement learning. The goal of this combination is to create self-learning software agents to establish winning strategies for maximizing long-term rewards. In DRL, values or policy functions are represented as deep neural networks, making it easy to apply related deep learning techniques, such as Deep Q_Network (DQN).

3.3.1 DEEP Q_NETWORK

DQN or Deep_Q Networks is one of the first successful algorithms that combine deep learning and reinforcement learning to learn approaches directly from high-dimensional raw data. DQN has accelerated the development of reinforcement learning and expanded its application scenarios by combining convolutional neural networks (CNNs) with Q_learning.

DQN was proposed by DeepMind in 2015 to integrate the benefits of deep learning into reinforcement learning [29]. Reinforcement learning focuses on training agents to take action at specific stages in the environment in order to maximize rewards.

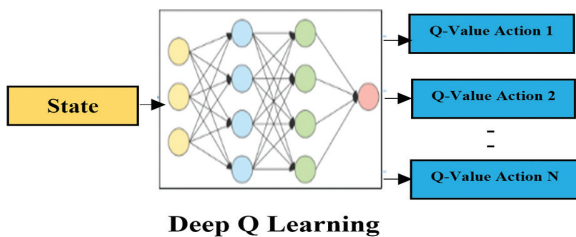


Fig. 3. The Deep Q-Network (DQN) algorithm

In this work, we have proposed a DQN network architecture to train and retrain our classification model.

3.4. PROPOSED METHODOLOGY

In this proposed study, a new and efficient skin cancer diagnosis model has been implemented for the precise classification and detection of malignant and benign skin cases.

Fig. 4 shows the complete workflow of our architecture consisting of two parts: Deep Learning and Reinforcement Learning.

The complete model consists of several steps starting from the input phase of applying the image preprocessing phase for analysis of the dermoscopic images before applying any feature extractor and classification methods [30]. Then, the first stage outputs are passed in the segmentation phase to determine the zone of cancerous skin and to effectively monitor the boundary areas of this zone. After that, the segmented image outputs are fed into Deep CNN Model to obtain the output probability.

In the final part, the result of the classification model is passed with a reinforcement learning environment with the DQN Algorithm to train and retrain our classification model up to find the best accuracy of classification task of skin cancer with the use of deep reinforcement learning.

The principal steps of our approach are:

1. Collecting the dermoscopic images for the dataset from skin cancer types images.
2. Applying image preprocessing techniques, such as reading, resizing images, cleaning images, and applying One Hot Encoding on the labels of the dataset.
3. Using Train_Test_Split to splitting the dataset into two sets: a training set and a test set.
4. Applying data segmentation using a watershed algorithm.
5. Extracting important features from images and classifying the dermoscopic images using the CNN network.
6. Reducing overfitting [31] and reducing error rates using dropout.
7. Training and retraining our classification model 'CNN' with reinforcement learning algorithm 'DQN'.
8. Choosing the optimal hyper-parameter.
9. Evaluating our model on the test dataset.

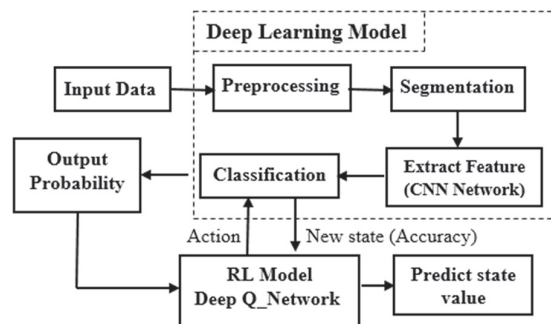


Fig. 4. The architecture of the proposed method.

Preprocessing: Image pre-processing is an important task that not only saves training time but also helps to significantly improve the next step by increasing the efficiency of the model. In this study, the preprocessing techniques used are:

- Collection of the dataset: we have collected 10015 dermatoscopic images of seven different classes of skin cancer types from the publicly available HAM10000 dataset.
- Import the necessary libraries.
- Reading and Resizing images: the images are resized to 28x28 to make the computation and training faster.
- Exploratory Dataset Analysis (EDA) to facilitate analysis of the data.
- Cleaning Dataset: remove duplicates, hair removal...
- Apply One-Hot Encoding for labels.
- Normalization: is a technique used to avoid the problems caused by the loss of contrast in the image.
- Splitting Dataset: The data were randomly split using the Train_Test_Split technique: 80% for training and 20% for testing.

Segmentation: After applying pre-processing techniques to the images to avoid all kinds of impurities, this image should be segmented, focusing on the area of interest to simplify the classification. Therefore, the next step is image segmentation. This is a technique used in image processing that divides an image into many parts, depending on the quality of the pixels in the image. Image segmentation involves recognizing and segmenting areas of interest based on how identical the colors and shapes are.

In our case, to extract the information about the skin cancer lesion with high confidence we have used the Watershed Algorithm is an efficient and successful segmentation technique in medical imaging. Because it's less sensitive to noise and is less computationally and calculation expensive. [32, 33]

Feature Extraction and Classification: CNN-based feature extraction and classification techniques are the most common method in image processing especially medical. here, next to image segmentation, the segmented image output passed to our CNN model.

We have proposed deep layers CNN, we increased the size and complexity of the model to give high accuracy and better efficiency. The structure of our CNN architecture proposed includes a sequence of convolutional layers, pooling layers, several activation functions, Batch Normalization, Dropout, and a Fully-Connected Layer "Softmax function" to obtain outputs of the probability of each class being present. The size of the convolution kernel was kept constant at all stages. In our case, to avoid the overfitting problem we used the average pooling layer with a small window size of 2*2.

Training and Evaluation: Recently, RL algorithms have been effectively and successfully combined with a Deep-NN [34]. This combination has been used to approximate RL functions using the Deep-NN model or when using RL to train Deep-NN.

In our case, we have applied the DQN algorithm with some fully connected layers and activation functions for train and retraining our deep CNN network until we found the best result with the optimal hyper-parameters. We followed a special process to train our model: Firstly, we selected random images for fed into the classification model to obtain the output probability (state). Then, we passed the result of the classification model in the RL model to predict state value.

Finally, in our RL environment, we selected an image with a height prediction value (Action) to train our CNN classification model to obtain the new state (Accuracy).

4. RESULTS

Our proposed system for the detection and classification of skin cancer was evaluated with the HAM10000 dataset which is described in the datasets section of the paper. This system uses deep reinforcement learning which combines the CNN and DQN algorithms thanks to their good performance. The proposed methodology was discussed in Section 3. And the performance of this last to classify the dermoscopic images was evaluated using the following parameter: accuracy.

We trained our proposed model with 80% of the training set by testing different optimizers such as SGD, RMSprop, Adam optimizer with a learning rate of 0.0001, and other hyper-parameters. After multiple tunings, we achieved the best result with 100 epochs and obtained an accuracy value of 80%. In light of this, the combination of deep learning (CNN) and reinforcement learning (DQN) is considered a more powerful and robust tool for classifying skin cancer. The idea is to take a decision based on many results obtained by training and retraining the CNN model.

Table 2. summarizes the accuracy obtained for training our proposed model.

Algorithm	Accuracy [%]	Optimizer	Learning rate
	69.57	Adam	1e-3
CNN+DQN	50.02	RMSProp	1e-5
	80	Adam	1e-4

We compare the performance of our method (CNN-DQN) and other methods while classifying skin cancer, as shown in Table 3.

Table 3. Comparison with the existing work.

Reference	Method	Accuracy [%]
[35]	DNN	76
[36]	CNN	79.45
[37]	EfficientNetB3	78
	Proposed Model	80

Our proposed model (CNN-DQN) performs better and outperforms other methods in terms of accuracy as seen in table 3, this explains the importance and the benefit of using the combination of deep learning and reinforcement learning for classification tasks.

5. DISCUSSION AND CONCLUSION

Skin cancer is a serious type of cancer, where increasing and affects many people every day. This cancer can be treated if it is detected in its early stages, but multiclass skin cancer diagnosis and classification is a tough undertaking. In addition, current clinical techniques used are few. For this reason, recently the IA algorithms were developed to support doctors to supplement clinical practices.

In this study, we proposed a new model as a solid and technical solution for the detection and classification of skin cancer combines with deep learning and reinforcement learning. This combination is important in making machines more intelligent in decision-making and obtaining high-performance accuracy in medical imaging and also it provides several advantages including optimal balancing between time efficiency and accuracy, dropping computational costs, and surging computing power. In our model, after preprocessing step, we passed with segmentation step to extract the affected area. Then, we used the deep convolutional neural network (CNN) for the classification part. And about the Reinforcement Learning part, we used the Deep Q_Learning algorithm to train and retrain our model until we find the best result. After extensive experiments, we obtained good results in comparison to some approaches presented in related work. we were able to achieve better classification accuracy by 80%, these results show the effectiveness of our approach.

We conclude to our model is among the best solution to make the decision in the healthcare domain especially medical image applications: in the case of skin cancer. Additionally, there is still a possibility for improvement if we supply additional data using specific and clean datasets and exploit the potential of deep reinforcement learning algorithms for the classification task.

6. REFERENCES:

- [1] Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer, <https://gco.iarc.fr/today> (accessed: 2021)
- [2] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. M. Swetter, H M. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, Vol. 542, No. 115, 2017, pp. 115-118.
- [3] A. A. Adegun, S. Viriri, "FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images", *IEEE Access*, Vol. 8, 2020, pp. 150377-150396.
- [4] "2020 Melanoma Skin Cancer Report Stemming the global epidemic", https://melanomapatients.org.au/wpcontent/uploads/2020/04/2020-campaign-report-GC-version MPA_1.pdf (accessed: 2023)
- [5] C. C. Darmawan, G. Jo, S. E Montenegro, Y. Kwak, L. Cheol, K. H. Cho, J-H. Mun, "Early detection of acral melanoma: a review of clinical, dermoscopic, histopathologic, and molecular characteristics", *Journal of the American Academy of Dermatology*, Vol. 81, 2019, pp. 805-812.
- [6] R. L. Siegel, K. D. Miller, A. Jemal, "Cancer statistics, 2018", *CA: A Cancer Journal for Clinicians*, Vol. 68, No. 1, 2018, pp. 7-30.
- [7] K. Das, C. J. Cockerell, A. Patil, P. Pietkiewicz, M. Giuliani, S. Grabbe, M. Goldust, "Machine Learning and Its Application in Skin Cancer", *International Journal of Environmental Research and Public Health*, Vol. 18, No. 24, 2021.
- [8] F. Mahmood, S. Bendayan, F. M. Ghazawi, I. V. Litvinov, "Editorial: The Emerging Role of Artificial Intelligence in Dermatology", *Frontiers in Medicine*, Vol. 8, 2021.
- [9] H. Rashid, M. A. Tanveer, H. A. Khan, "Skin Lesion Classification Using GAN Based Data Augmentation", *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Berlin, Germany, 23-27 July 2019*, pp. 916-919.
- [10] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, R. M. Summers, "A Bottom-Up Approach for Pancreas Segmentation Using Cascaded Super-pixels and (Deep) Image Patch Labeling", *IEEE Transactions on Image Processing*, Vol. 26, 2017, pp. 386-399.
- [11] M. Bakator, D. Radosav, "Deep Learning and Medical Diagnosis: A Review of Literature", *Multimodal Technologies and Interaction*, Vol. 2, 2018.
- [12] R. S. Sutton, A. G. Barto, "Introduction to Reinforcement Learning", MIT Press, 2015.
- [13] S. Magalhães Barros Netto, V. Rodrigues Coelho Leite, A. Silva, A. Paiva, "Application on Reinforcement Learning for Diagnosis Based on Medical Image", *Reinforcement Learning*, Intech Open, 2008, p. 424.
- [14] Z. Liu, H. Yu, T. Wu, "Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things", *Future Generation Computer Systems*, Vol. 97, 2019, pp. 1-9.
- [15] A. T. Simin, S. M. G. Baygi, A. Noori, "Cancer Diagnosis Based on Combination of Artificial Neural Networks

- and Reinforcement Learning", Proceedings of the 6th Iranian Conference on Signal Processing and Intelligent Systems, 2020.
- [16] M. K. Monika, N. A. Vignesh, C. U. Kumari, M.N.V.S.S. Kumar, E. L. Lydia, "Skin cancer detection and classification using machine learning", *Materials Today: Proceedings*, Vol. 33, 2020, pp. 4266-4270.
- [17] M. Gaana, S. Gupta, N. S. Ramaiah, "Diagnosis of Skin Cancer Melanoma using Machine Learning", *SSRN*, 2019, <https://ssrn.com/abstract=3358134> (accessed: 2023)
- [18] M. Q. Hatem, "Skin lesion classification system using a K-nearest neighbor algorithm", *Visual Computing for Industry, Biomedicine, and Art*, Vol. 5, No. 7, 2022.
- [19] S. Ali, S. Miah, J. Haque, M. Rahman, K. Islam, "An enhanced technique of skin cancer classification using a deep convolutional neural network with transfer learning models", *Machine Learning with Applications*, Vol. 5, 2021, p. 100036.
- [20] L. Moataz, G. I. Salama, M. H. Abd Elazeem, "Skin Cancer Diseases Classification using Deep Convolutional Neural Network with Transfer Learning Model", *Proceedings of the 6th International Conference on Advanced Technology and Applied Sciences*, Cairo, Egypt, 2021.
- [21] B. Samia, M. Boudjelal, O. L  zoray, "Skin lesion classification using convolutional neural networks based on Multi-Features Extraction", *Proceedings of the 19th International Conference on Computer Analysis of Images and Patterns*, 2021.
- [22] R. Kaur, H. G. Hosseini, R. Sinha, M. Lind  n, "Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images", *Sensors*, Vol. 22, No. 3, 2022.
- [23] M. Nawaz, Z. Mehmood, T. Nazir, R. A. Naqvi, A. Rehman, M. Iqbal, T. Saba, "Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering", *Microscopy Research and Technique* Vol. 85, 2021, pp. 339-351.
- [24] Y. N. Fu'adah, N. K. C. Pratiwi, M. A. Pramudito, N. Ibrahim, "Convolutional Neural Network for Automatic Skin Cancer Classification System", *Proceedings of the Materials Science and Engineering, International Conference in Engineering, Technology and Innovative Researches*, Purbalingga, Indonesia, Vol. 982, 2020.
- [25] P. Tschandl, C. Rosendahl, H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", *Scientific Data*, Vol. 5, No. 180161, 2018.
- [26] A. Krizhevsky, I. Sutskever, G. Hinton, "Imagenet classification with deep convolutional neural networks", *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, 2012.
- [27] Y. LeCun, "LeNet-5, Convolutional Neural Networks", <http://yann.lecun.com/exdb/lenet> (accessed: 2023)
- [28] R. S. Sutton, A. G. Barto, "Introduction to Reinforcement Learning", *IEEE Transactions on Neural Networks*, Vol. 9, 1998.
- [29] V. Mnih et al. "Human-level control through deep reinforcement learning", *Nature*, Vol. 518, 2015.
- [30] C. Fan, M. Chen, X. Wang, J. Wang, B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery from Building Operational Data", *Frontiers in Energy Research*, Vol. 9, 2021.
- [31] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, 2017. pp. 84-90.
- [32] V. Shanthi, G. Sridevi, R. Charanya, J. Josphin Mary, "Watershed Algorithm in Multichannel for Skin Lesion Segmentation", *European Journal of Molecular & Clinical Medicine*, Vol. 7, 2020, pp. 1374-1378.
- [33] U. B. Ansari, T. Sarode, "Skin Cancer Detection using Image Processing", *International Research Journal of Engineering and Technology*, Vol. 4, No. 4, 2017.
- [34] Y. Li, "Deep reinforcement learning: An overview", *arXiv:1701.07274*, 2018.
- [35] M. A. A. Milton, "Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge", *arXiv:1901.10802*, 2019.
- [36] N. Rezaoana, M. S. Hossain, K. Andersson, "Detection and Classification of Skin Cancer by Using a Parallel CNN Model", *Proceedings of the IEEE International Women in Engineering Conference on Electrical and Computer Engineering*, Bhubaneswar, India, 26-27 December 2020.
- [37] I. U. Khan et al. "Remote Diagnosis and Triaging Model for Skin Cancer Using EfficientNet and Extreme Gradient Boosting", *Complexity*, Vol. 2021, 2021, p. 5591614.

Real-World Anomaly Detection in Video Using Spatio-Temporal Features Analysis for Weakly Labelled Data with Auto Label Generation

Original Scientific Paper

Rikin J. Nayak

VT Patel Dept of E & C Engg, Chandubhai S Patel Institute of Technology,
Charotar University of Science and Technology, Changa, Ta-Petlad, Anand, Gujarat 388421, India
rikinnayak@gmail.com, 16drec006@charusat.edu.in

Jitendra P. Chaudhari

Charusat Space Research and Technology Center, VT Patel Dept of E & C Engg,
Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa,
Ta-Petlad, Anand, Gujarat 388421, India jitendrachaudhari.ec@charusat.ac.in

Abstract – Detecting anomalies in videos is a complex task due to diverse content, noisy labeling, and a lack of frame-level labeling. To address these challenges in weakly labeled datasets, we propose a novel custom loss function in conjunction with the multi-instance learning (MIL) algorithm. Our approach utilizes the UCF Crime and ShanghaiTech datasets for anomaly detection. The UCF Crime dataset includes labeled videos depicting a range of incidents such as explosions, assaults, and burglaries, while the ShanghaiTech dataset is one of the largest anomaly datasets, with over 400 video clips featuring three different scenes and 130 abnormal events. We generated pseudo labels for videos using the MIL technique to detect frame-level anomalies from video-level annotations, and to train the network to distinguish between normal and abnormal classes. We conducted extensive experiments on the UCF Crime dataset using C3D and I3D features to test our model's performance. For the ShanghaiTech dataset, we used I3D features for training and testing. Our results show that with I3D features, we achieve an 84.6% frame-level AUC score for the UCF Crime dataset and a 92.27% frame-level AUC score for the ShanghaiTech dataset, which are comparable to other methods used for similar datasets.

Keywords: anomaly detection, spatio-temporal analysis, 3d convolutional neural network, multi-instance learning

1. INTRODUCTION

Anomaly detection, or the identification and classification of data patterns that deviate from normal patterns, is a crucial aspect of intelligent visual surveillance systems. The deployment of CCTV cameras has become widespread and more affordable, which has resulted in increased research attention on video-based anomaly detection. Deployment of CCTV cameras is particularly important for ensuring security in public areas such as railway stations, hospitals, and military bases. With the increasing availability of powerful computing resources, Artificial Intelligence and Deep learning have been integrated into smart video surveillance systems to efficiently process and analyze vast amounts of video data. In this paper, we employ a multi-instance learning technique to address the challenges of anomaly detection in videos, using a custom loss function under weakly supervised learning. We assess the effectiveness of our approach on two different datasets, we compare various feature extraction techniques and performance metrics.

Artificial Intelligence and Deep Learning have significantly enhanced smart video surveillance systems. The effectiveness of these approaches relies on substantial processing power, large datasets, and advanced resources, which have become increasingly accessible due to powerful GPUs and high-RAM systems. Although convolutional neural networks (CNNs) excel at processing spatial information in images, they face limitations when analyzing temporal information in videos. Recurrent neural networks, such as Long Short-Term Memory (LSTM) networks, can address this challenge by modeling sequence information in video data, where each frame depends on its predecessors.

Various studies have explored anomaly detection [1-5], employing two main approaches based on the availability of labeled data. The traditional method, suited for situations where labeled data is unavailable, trains the model using known normal data. Alternatively, if labeled data is available, it can be used to train the model and predict abnormal classes for future test data. According to D. Elliott (2010), after 12 hours, a sin-

gle person can miss up to 80% of the activity between two cameras. This highlights the importance of effective anomaly detection systems. Deep learning outperforms other methods when the available dataset is large [6]. Anomaly detection or outlier detection is useful in various applications, including detecting illegal traffic flow [7], retinal damage [8], and IoT big-data anomaly detection [9]. However, deep learning-based methods often face difficulties in anomaly detection because of the complex structure of the data and the narrow boundary between normal and abnormal data.

In this paper, we make the following contributions:

1. We utilize multi-instance learning technique with auto label generation loss to tackle the challenge of anomaly detection in videos, particularly when video-level labels are available, but anomalies occur at the frame level.
2. We introduce a custom loss function for use in weakly supervised learning, designed to enable more effective extraction of discriminative features and thereby improve anomaly detection performance.
3. We incorporate a mean squared error function on auto-generated labels, which aids in separating interclass features and increasing intraclass feature closeness.
4. Our experiments, conducted without sparsity and temporal smoothness constraints, show that our proposed model is robust and effective. We evaluate our model on two benchmark datasets, UCF Crime and Shanghaitech, using various feature extraction techniques and comparing proposed loss functions in different environments.
5. We demonstrate that the I3D feature extractor outperforms the C3D feature extractor in our experiments, and we assess the model's performance using the area under the curve (AUC) metric.

The paper is organized as follows: Section 2 reviews related work and presents the problem statement, our proposed approach discussed in Section 3, Section 4 discusses experimental results, and Section 5 concludes the paper.

2. RELATED WORK

Deep learning has proven to be a superior approach compared to traditional machine learning in several areas, particularly in image and video processing. Despite deep learning's superiority in various areas, detecting anomalies in image and video processing remains a challenging task, with many researchers making significant contributions to this field [10-14]. In [10], particle trajectories were utilized to model normal motion, and deviations from the norm were defined as anomalies. The author in [15] provides an in-depth analysis of deep anomaly detection in the medical domain. Researchers have also explored violence and aggression detection [16-18].

Feature learning is a conventional approach for inferring normality from data. However, due to difficulties associated with tracking objects in videos, many researchers have employed alternative methods such as motion pattern analysis using a histogram-based method [19], kernel density estimation methods [20], social force models [21], context-driven methods [22], and hidden Markov models [23]. These techniques offer different ways to address the difficulties of understanding motion and detecting deviations from normal patterns. During the testing phase, videos with lower probability are classified as anomalies, while normal videos are used for training. In [24], researchers focused on the problem of online detection of unusual events in videos using dynamic sparse coding. The main idea is that sparse representation can help us learn about normal behaviour in videos, which can then be used to detect unusual or abnormal events. Developing a video action classification model using deep learning has been proposed in [25]. However, video classification is more challenging than deep learning-based image classification due to the difficulty of obtaining annotations for training the model and the extensive efforts required to generate frame-level labels. To address the challenges posed by weakly labeled datasets, researchers have explored various approaches, as discussed in [26-29].

The RTFM(Robust Temporal Feature Magnitude learning) method [27] enhances detection by training a specialized function to recognize rare events and consider their timing, resulting in better accuracy and efficiency for detecting subtle anomalies. The MIST framework [28] focuses on using video-level annotations to refine important features, making the anomaly detection process more effective. Furthermore, the authors in [29] introduced the LAD database, a comprehensive collection of video sequences for anomaly detection, along with a multi-task deep neural network that leverages spatiotemporal features, achieving superior performance compared to existing methods in the field.

The author in [26] utilized a multi-instance learning (MIL) model to address the issue of weakly labeled datasets. Similar approaches have been employed for detecting anomalies, as discussed in [30-33]. The author in [31] proposes the Anomaly Regression Net (ARNet) framework for video anomaly detection, which only requires video-level labels in training and uses multiple-instance learning loss and centre loss for discriminative features. [32] proposes a weakly supervised deep temporal encoding-decoding solution using multiple instance learning for anomaly detection in surveillance videos and employs a new smoother loss function. [33] focuses on reducing false alarms in abnormal activity detection using 3D ResNet and deep multiple instance learning with a new ranking loss function, achieving the best performance on the UCF-Crime benchmark dataset. All three papers present novel approaches for video anomaly detection and achieve advanced results on challenging benchmark datasets.

Detecting anomalies with accuracy is a challenging task, primarily due to its subjective nature, which varies based on location and individual perspectives. Researchers have approached anomaly detection as a means of identifying low-probability patterns, as evident in studies conducted by [34-36]. In this research, we address the problem of anomaly detection as a regression issue and propose a customized loss function, coupled with multi-instance learning techniques.

Our proposed loss function aims to increase the gap between the normal and abnormal frames while minimizing computational complexity. This is achieved by removing the sparsity and temporal smoothness constraints typically present in similar techniques. The proposed methodology section will detail our approach to addressing the challenge of detecting anomalies with high accuracy.

2.1. PROBLEM STATEMENT

Our research tackles the challenge of frame-level anomaly detection in videos using the UCF Crime and Shanghai tech datasets. These datasets provide anomaly labels at the video level, complicating frame-level detection. To address this, we employ multi-instance learning (MIL) and split the dataset into two parts: one with normal frames and another with a mixture of normal and abnormal frames grouped under a single anomaly class. Our aim is to effectively detect anomalies at the frame level by utilizing MIL and a custom loss function that minimizes false anomaly detections. We will detail our techniques, their application to the datasets, and our experimental results in the subsequent sections. By enhancing frame-level anomaly detection, our research contributes to the field of video surveillance and has potential applications in security systems and public safety measures.

3. PROPOSED METHOD

This section of the paper aims to define the problem of anomaly detection in video, describe the feature

extraction method, and provide a detailed description of the proposed loss function. To detect anomalies in the video, we utilize the UCF Crime and Shanghai tech datasets, which contain a range of videos of different lengths categorized as normal, explosive, burglary, fighting, and arrest. Similar to [26], anomaly detection is treated as a regression problem, where a sequence of frames serves as the input and an anomaly score between 0 and 1 is the output for each frame.

In this work, we present a deep learning method-based approach for detecting anomalies. We begin by converting the input video into a fixed-size array and then extract features using both three-dimensional convolutional features [37] and inflated three-dimensional (I3D) features [38]. Each video is then segmented into a fixed number of non-overlapping temporal segments, and each segment is treated as a "bag" instance for feature extraction. We extract 3D convolutional and I3D features from each video segment.

We utilized two types of pre-processed video data, namely C3D and I3D features, to extract features for our model. These models were chosen due to their efficiency in learning spatiotemporal features, which are crucial for further processing. C3D features consist of two-stream pre-processed video data with a feature dimension of 4096. On the other hand, I3D features are composed of RGB and optical-flow features, with a feature dimension of 2048 for each. During the training process, we concatenated the RGB and optical flow features to create a unified input. To visualize our proposed approach, we have included a diagram of the model with the custom loss function in Fig. 1.

The loss function of the support vector machine model is

$$L(w) = \frac{1}{k} \sum_{i=1}^k \max(0, 1 - y_i | w^T x_i + b|) + \lambda ||w||_2^2 \quad (1)$$

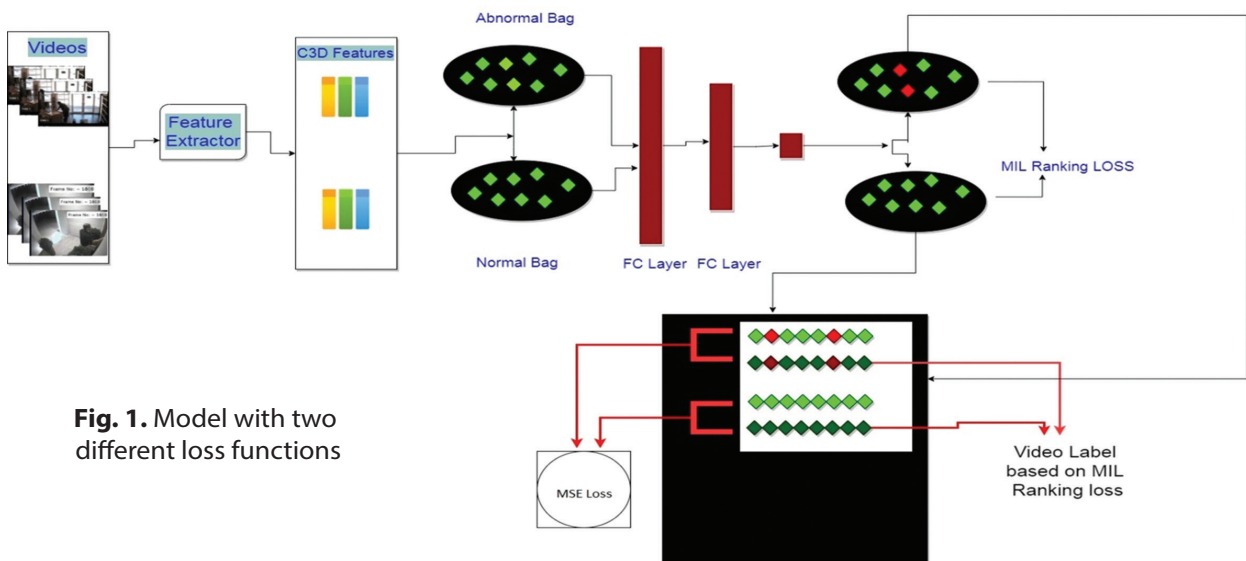


Fig. 1. Model with two different loss functions

where w is the weight vector, b is the bias term, and T denotes the transpose of the weight vector. The loss is accumulated over all training data points and is often combined with a regularization term to prevent overfitting. The loss function used in the model has two components: the hinge loss and the regularization term. During training, the learning parameter w is adjusted to minimize the hinge loss, which generates a positive loss for incorrectly classified features. In this supervised learning approach, the labels Y_i and features x_i are used along with the bias b to determine the loss. However, as the video frames lack annotation, this approach is not applicable. To address this issue, the MIL approach was adopted, as discussed in [26]. Under this approach, each video is divided into a bag, with the positive bag containing both normal and abnormal frames and the negative bag containing only normal frames. Similar to [26, 39], the maximum $(w^T x_i + b)$ is considered for both types of bags. This approach allows the model to learn to identify abnormal frames without the need for individual frame annotations, which will modify equation 1 to

$$L(w) = \frac{1}{k} \sum_{i=1}^k \max(0, 1 - y_i \max_{i \text{ from bag}} (|w^T x_i + b|)) + \lambda \|w\|_2^2 \quad (2)$$

Here, Y_i is the bag-level label. where y is the true label of the bag (+1 for positive bags and -1 for negative bags), $f(x)$ is the decision function of the SVM, B is the set of instances in the bag, and $\max_{i \text{ from bag}} (|w^T x_i + b|)$ is the maximum predicted score for any instance in the bag.

Our proposed loss function aims to maximize the distance between positive and negative bags, with only the maximum distance feature considered for each bag. The selection of the maximum feature is based on the assumption that each abnormal bag should contain at least one abnormal instance, while a normal bag should only contain normal instances. Building on this approach, we developed a custom loss function that combines multi-instance learning with the residual difference between actual and predicted labels to train the network. In this case, the actual label is determined through maximum selection in the MIL process. This can be explained by equation 3:

$$L_{MIL} = \frac{1}{kN} \sum_{kN}^{Max_{i \in Ni}} (y_i * (w^t x_i + b)) + \gamma - \frac{1}{KAN} \sum_{kAN}^{Max_{i \in Abi}} (y_i * (w^t x_i + b)) \quad (3)$$

Here, the first term represents the average sum of the maximum distanced feature from each normal video, and the second term represents the average sum of the maximum distanced feature from each abnormal video, and the third term represents the hyperparameter. The objective of this loss function is to maximize the difference between the abnormal and normal features, as represented by the first two terms of the equation. To further refine this approach, we introduce a custom loss function that combines multi-instance learning

with the residual difference between the actual and predicted labels.

The following equation explains how pseudo-labels are generated for each instance:

$$L_{MSE} = \frac{1}{k} \sum_k (Y_{AUTO} - Y_{MIL})^2 \quad (4)$$

Here, Y_{AUTO} is a label generated based on the distance of the feature from the line measured by L_{MIL} . Y_{AUTO} assigns a label of 1 (abnormal) to an instance if the maximum absolute value of the weighted sum for all instances in the bag is greater than a certain threshold. Otherwise, it assigns a label of 0 (normal). This method helps identify the most representative instances within each bag, which, in turn, assists in the training of the network to maximize the difference between normal and abnormal features. Y_{AUTO} is calculated as follows:

$$Y_{AUTO} = 1 \text{ if } \max_{i \text{ from bag}} (|w^T x_i + b|) \text{ else } 0 \quad (5)$$

Y_{MIL} is the actual distance calculated for each feature in the bag. The final loss function is the sum of equations (3) and (4).

$$L_{MIL_MSE} = L_{MIL} + L_{MSE} \quad (6)$$

Combining multi-instance learning with the residual difference The custom loss function incorporates both the multi-instance learning component (L_{MIL}) and the residual difference between actual labels and pseudo-labels (L_{MSE}). This combination allows the model to better learn the relationship between the features and the labels, resulting in improved anomaly detection.

Using Y_{AUTO} as a pseudo-label helps the model learn better decision boundaries by leveraging the information from the most representative instances. This aids in training the model to effectively distinguish between normal and abnormal instances, improving its overall anomaly detection capability.

4. EXPERIMENTAL RESULTS

This section describes the use of C3D and I3D as feature extractors for video anomaly detection on the UCF Crime dataset and the ShanghaiTech dataset. C3D is a neural network that extracts spatiotemporal features from videos, while I3D is a modified version of C3D that achieves advance results in video recognition tasks. The proposed approach extracts features using pre-trained C3D and I3D networks and uses a one-class SVM classifier with a custom loss for anomaly detection. The one-class SVM classifier is a popular choice for anomaly detection as it is designed to distinguish between normal and abnormal instances. The experimental results show that I3D outperforms C3D in all evaluation metrics, and the system's performance improves with an increase in the number of frames used in feature extraction. The proposed approach achieves competitive results compared to state-of-the-art methods on the UCF Crime dataset and the ShanghaiTech dataset.

The UCF Crime dataset and the ShanghaTech dataset are both challenging and widely used benchmark datasets for video anomaly detection. The UCF Crime dataset consists of 1,900 real-world surveillance videos that encompass various crime types, such as theft, robbery, vandalism, and fights. This diverse dataset poses a challenge for models to accurately detect and classify different types of anomalous behaviours in realistic settings. On the other hand, the ShanghaTech dataset contains 437 high-resolution surveillance videos from diverse environments like streets, parks, and commercial areas, featuring anomalies such as jaywalking, loitering, and illegal parking. Its difficulty arises from the high variability in video content, camera angles, and lighting conditions, making it a robust dataset for evaluating video anomaly detection model performance across different scenarios.

4.1. C3D NETWORK

This section introduces a video anomaly detection approach utilizing C3D features extracted from the UCF Crime dataset, as outlined in [26]. The C3D features capture both the appearance and dynamics of moving objects for video action recognition. Each video is segmented into non-overlapping fixed-size segments to create a 4096x32 feature matrix. A neural network having four fully interconnected layers with 256, 64, and 16 neurons and a single output neurons is employed, using an Adagrad optimizer and a learning rate of 0.01. The performance is assessed by the area under the receiver operating characteristic (AUC-ROC) curve, enabling fair comparisons. This approach computes the ROC curve based on the frame-level anomaly score.

4.2. I3D NETWORK

This experiment adopts the Inflated 3D (I3D) model, pre-trained on the Kinetics dataset, as the feature extraction network. The I3D network output for each video includes RGB and optical flow features, which are concatenated, producing a 2048x32 feature output size. A four-layer fully connected neural network with 128, 32, and 16 units and a single output layer is used. Training is conducted with the Adagrad optimizer and a 0.01 learning rate. Tables 1 and 2 display the results of our custom loss function.

Table 1 highlights the effectiveness of incorporating I3D features into the model for video anomaly detection. The I3D features-based approach achieves an AUC score of 84.66, surpassing other methods in the comparison, thus demonstrating its superiority. Tests were also conducted using C3D features and I3D with only RGB features. Table 2 summarizes the corresponding AUC, F1, and EER scores, providing insights into the performance of different feature sets in video anomaly detection and emphasizing the advantages of I3D features.

Our experiments, conducted using the open-source code by Sultani et al. [26], are based on established research and methods. A confusion matrix in Table 3 adds

context and understanding to our findings, detailing the rates of true and false predictions, enabling readers to evaluate the model's effectiveness in detecting video anomalies comprehensively.

Overall, our results in Tables 1, 2, and 3 strongly support I3D features for video anomaly detection. The high AUC score, F1 score, and EER emphasize the effectiveness of our approach compared to others. Incorporating I3D features yields the best performance, as indicated by the highest AUC score. These findings have important implications for future research. Fig. 2 and 3 display results for various test dataset videos. Fig. 3 illustrates the anomaly score graph for abnormal frames when the model generates higher scores compared to normal frames. This figure presents the results for two specific video instances: a) Stealing079_x264 and b) Stealing047_x264. The visual representation in Fig. 3 provides insights into the model's ability to accurately detect and distinguish abnormal behavior, such as theft, from regular activities, further showcasing the effectiveness of the model in video anomaly detection tasks.

Table 1. AUC Score of Comparison on UCF Crime dataset with Various methods

Method	Features	AUC (%)
Hasan et al [35]	C3D RGB	50.6
Lu et al [40]	C3D RGB	65.51
Sultani et al. [26]	C3D RGB	75.41
MIST [28]	C3D RGB	81.40
	I3D RGB	82.30
Zhang et al.[41]	C3D RGB	78.66
J. Zhong et al [42]	C3D	81.08
	TSNRGB	82.12
	TSNOptical Flow	78.08
Proposed	C3D RGB	76.004
	I3D RGB	81.55
	I3D RGB + Optical Flow	84.66

Table 2. AUC Score of Comparison on UCF Crime dataset : C3D vs I3D

Features	AUC (%)	F1 Score	EER
C3D	76.00	25.61	30.75
I3D RGB + Optical Flow	84.66	35.63	22.59

Table 3. Confusion Matrix

	Predicted: Normal	Predicted: Abnormal
Actual: Normal	82.85 % (851346)	17.14 % (176131)
Actual: Abnormal	33.03 % (27860)	66.96 % (56471)

The model accurately predicts vandalism and stealing anomalies but doesn't generate an alert for normal videos. Due to varying conditions and challenges, the model isn't flawless at anomaly detection. In Fig. 4, a) the model occasionally fails to generate alerts when anomalies are present, and b) the model generates false alerts in the absence of visible anomalies. Generally, our network produces higher scores for abnormal video segments.

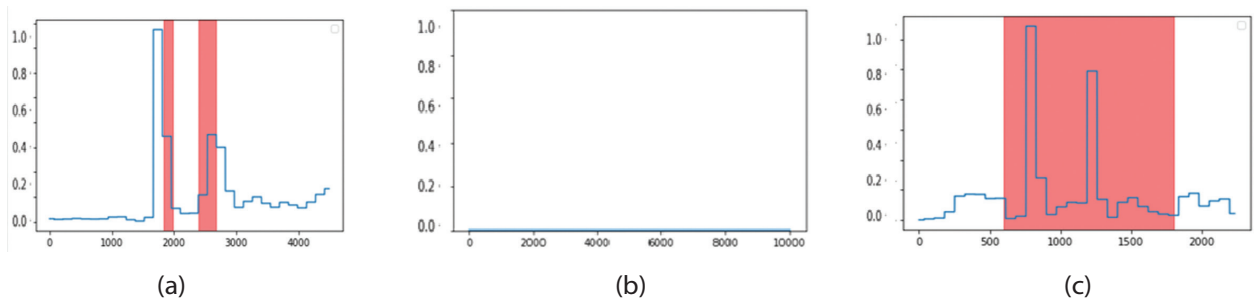


Fig. 2. Visualization of testing Results for a). Vandalism028, b). Normal877, c). Stealing059 Red portion shows actual abnormal frames Blue line shows anomaly score for given frames

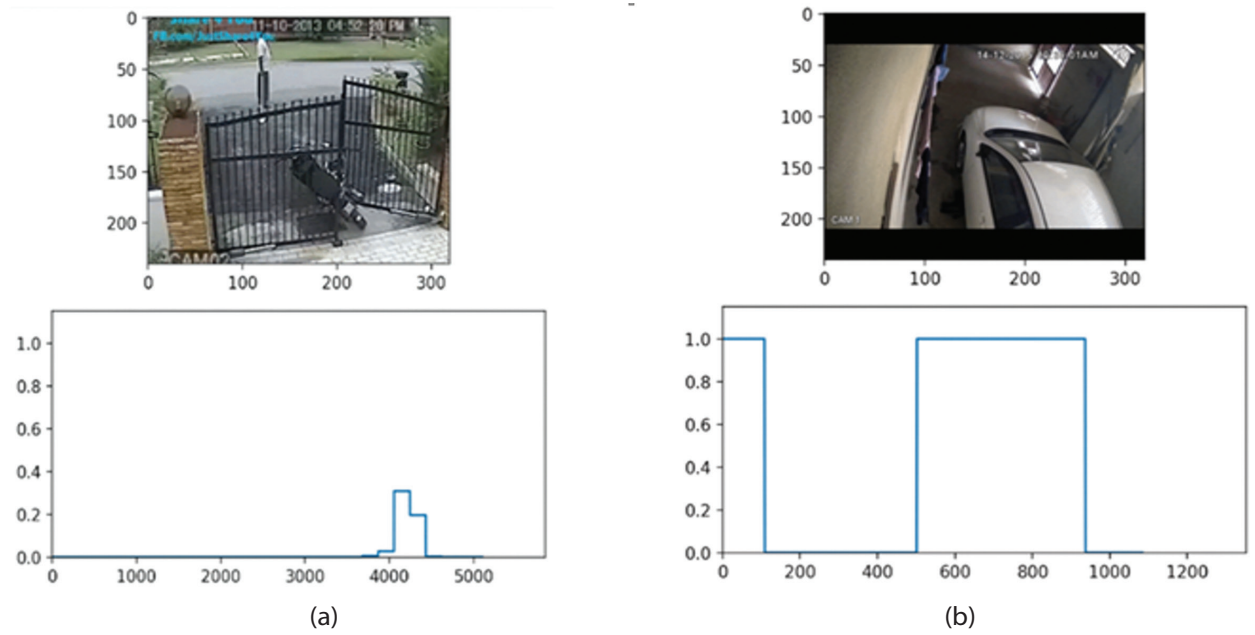


Fig. 3. Results of anomaly score graph for abnormal frames when model generates higher score compare to normal frames a) Stealing079_x264 b) Stealing047_x264

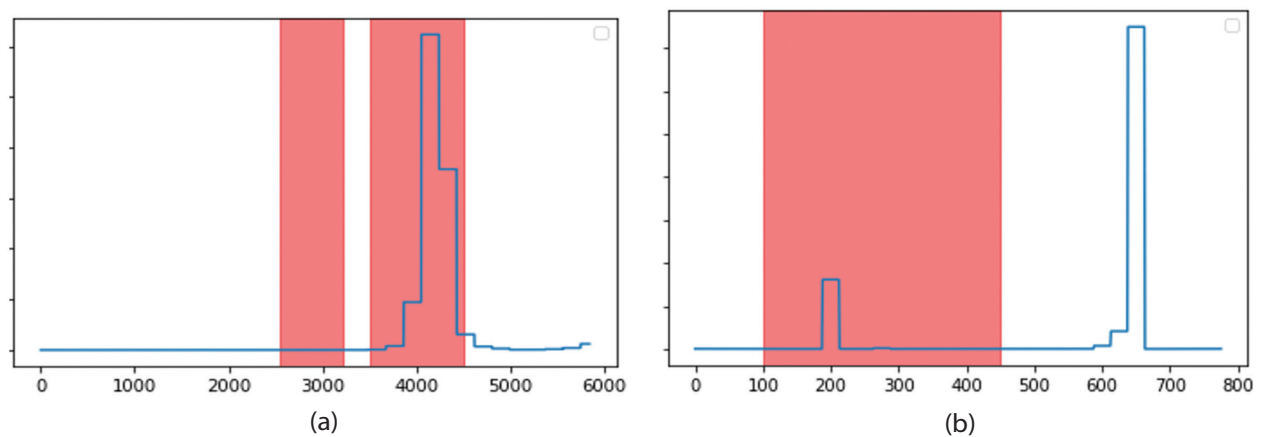


Fig. 4. Wrong prediction Results a) Stealing079_x264 here Model is not predicting anomaly for first window b) Explosion027_x264 Model is generating wrong alarm even anomaly is not there in video

The ShanghaTech dataset, a challenging benchmark dataset, is employed to further evaluate our model's performance. In this case, we consolidate all normal videos into a single normal class and all abnormal videos into a single abnormal class. Following the dataset split suggested in [42], we facilitate binary categorization. The dataset consists of 238 training videos and

199 test videos. Table 4 presents the AUC score results using l3D features for the ShanghaTech dataset. Our proposed model outperforms the other methods listed in Table 4, achieving an impressive AUC score of 92.27. This performance on the ShanghaTech dataset, known for its difficulty, further validates the effectiveness of our model in anomaly detection tasks.

Table 4. AUC Score of Quantitative Comparison on Shanghai Tech dataset

Method	Features	AUC(%)
Zhong et al.	TSNRGB	84.44
Zhong et al.	TSNOptical-Flow	84.13
Sultani et al [26]	C3D	86.30
AR-NET	I3D conc (RGB + Optical Flow)	91.24
Proposed (I3D RGB + Optical Flow)	I3D conc(RGB + Optical Flow)	92.27

5. CONCLUSION

This article presents a novel approach for detecting anomalies in videos using multi-instance learning and a custom dynamic loss function called LMIL_MSE. The loss function is calculated using the mean square error and is influenced by how well interclass features are separated. Pseudo-label generation is also used to improve the quality of the training data. Our experiments on the UCF Crime and Shanghai Tech datasets show that our approach outperforms previous methods in detecting video anomalies. To extract features from the video data, we used both C3D and I3D networks. We found that the I3D features yielded the highest AUC score in our experiments. Additionally, we utilized a multi-instance learning approach to improve the detection of anomalies in the video data.

Our research offers a promising solution to the challenging problem of detecting video anomalies. By achieving better results than previous methods, our approach has the potential to enhance the accuracy and reliability of video anomaly detection systems in real-world applications.

6. REFERENCES:

- [1]. C. C. Aggarwal, "An introduction to outlier analysis", *Outlier analysis*, Springer Cham., 2017, pp. 1-34.
- [2]. A. Deng, B. Hooi. "Graph neural network-based anomaly detection in multivariate time series", *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, 2-9 February 2021, pp. 4027-4035.
- [3]. A. Boukerche, L. Zheng, O. Alfandi, "Outlier detection: Methods, models, and classification", *ACM Computing Surveys*, Vol. 53, No. 3, 2020, pp. 1-37.
- [4]. A. Zimek, E. Schubert, H.-P. Kriegel. "A survey on unsupervised outlier detection in high-dimensional numerical data", *Statistical Analysis and Data Mining*, Vol. 5, No. 5, 2012, pp. 363-387.
- [5]. R. Chalapathy, S. Chawla, "Deep learning for anomaly detection: A survey", *arXiv:1901.03407*, 2019.
- [6]. A. Bahnsen, "Building AI Applications Using Deep Learning", <https://albahnsen.com/2017/06/06/building-ai-applications-using-deep-learning/> (accessed: 2023)
- [7]. X. Xie, C. Wang, S. Chen, G. Shi, Z. Zhao, "Real-time illegal parking detection system based on deep learning", *Proceedings of the International Conference on Deep Learning Technologies*, 2017, pp. 23-27.
- [8]. T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery", *Proceedings of Information Processing in Medical Imaging: 25th International Conference*, Boone, NC, USA, 25-30 June 2017, pp. 146-157.
- [9]. M. Mohammadi, A. Al-Fuqaha, S. Sorour, M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey", *IEEE Communications Surveys & Tutorials*, Vol. 20, No. 4, 2018, pp. 2923-2960.
- [10]. S. Wu, B. E. Moore, M. Shah, "Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 2010, pp. 2054-2060.
- [11]. D. Xu, E. Ricci, Y. Yan, J. Song, N. Sebe. "Learning deep representations of appearance and motion for anomalous event detection", *arXiv:1510.01553*, 2015.
- [12]. B. Antić, B. Ommer, "Video parsing for abnormality detection", *Proceedings of the International Conference on Computer Vision*, Barcelona, Spain, 6-13 November 2011, pp. 2415-2422.
- [13]. V. Singh, S. Singh, P. Gupta, "Real-Time Anomaly Recognition Through CCTV Using Neural Networks", *Procedia Computer Science*, Vol. 173, 2020, pp. 254-263.
- [14]. Md Sharif, L. Jiao, C. W. Omlin, "Deep Crowd Anomaly Detection: State-of-the-Art, Challenges, and Future Research Directions", *arXiv:2210.13927*, 2022.
- [15]. G. Litjens et al. "A survey on deep learning in medical image analysis", *Medical Image Analysis*, Vol. 42, 2017, pp. 60-88.

- [16]. S. A. A. Akash, R. S. S. Moorthy, K. Esha, N. Nathiya, "Human Violence Detection Using Deep Learning Techniques", *Journal of Physics: Conference Series*, Vol. 2318, No. 1, 2022, p. 012003.
- [17]. H. Gupta, S. T. Ali, "Violence Detection using Deep Learning Techniques", *Proceedings of the International Conference on Emerging Techniques in Computational Intelligence*, Hyderabad, India, 2022, pp. 121-124.
- [18]. S. Sadiq, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, B.-W. On, "Aggression detection through deep neural model on Twitter", *Future Generation Computer Systems*, Vol. 114, 2021, pp. 120-129.
- [19]. X. Cui, Q. Liu, M. Gao, D. N. Metaxas, "Abnormal detection using interaction energy potentials", *Proceedings of the CVPR 2011*, Colorado Springs, CO, USA, 20-25 June 2011, pp. 3161-3167.
- [20]. G.-h. Ji, X.-h. Zhang, X.-y. Cheng, "Pedestrians Detection Based on the Integration of Human Features and Kernel Density Estimation", *IOP Conference Series: Materials Science and Engineering*, Vol. 490, No. 4, 2019, p. 042027.
- [21]. X. Yang et al. "Deep social force network for anomaly event detection", *IET Image Processing*, Vol. 15, No. 14, 2021, pp. 3441-3453.
- [22]. Y. Zhu, N. M. Nayak, A. K. Roy-Chowdhury, "Context-aware activity recognition and anomaly detection in video", *IEEE Journal of Selected Topics in Signal Processing*, Vol. 7, No. 1, 2012, pp. 91-101.
- [23]. L. Kratz, K. NishiNo, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 20-25 June 2009, pp. 1446-1453.
- [24]. B. Zhao, L. Fei-Fei, E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding", *Proceedings of the CVPR 2011*, Colorado Springs, CO, USA, 20-25 June 2011, pp. 3313-3320.
- [25]. A. Karpathy, G. Toderici, S. Shetty, Thomas Leung, R. Sukthankar, L. Fei-Fei, "Large-scale video classification with convolutional neural networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23-28 June 2014, pp. 1725-1732.
- [26]. W. Sultani, C. Chen, M. Shah, "Real-world anomaly detection in surveillance videos", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18-23 June 2018, pp. 6479-6488.
- [27]. Y. Tian, G. Pang, Y. Chen, R. Singh, J. W. Verjans, G. Carneiro, "Weakly-supervised video anomaly detection with robust temporal feature magnitude learning", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, 10-17 October 2021, pp. 4975-4986.
- [28]. J.-C. Feng, F.-T. Hong, W.-S. Zheng, "MIST: Multiple instance self-training framework for video anomaly detection", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 20-25 June 2021, pp. 14009-14018.
- [29]. B. Wan, W. Jiang, Y. Fang, Z. Luo, G. Ding, "Anomaly detection in video sequences: A benchmark and computational model", *IET Image Processing*, Vol. 15, No. 14, 2021, pp. 3454-3465.
- [30]. B. Wan, Y. Fang, X. Xia, J. Mei, "Weakly supervised video anomaly detection via center-guided discriminative learning", *Proceedings of the IEEE International Conference on Multimedia and Expo*, London, UK, 6-10 July 2020, pp. 1-6.
- [31]. A. M. Kamoona, A. K. Gostar, A. Bab-Hadiashar, R. Hoseinnezhad. "Multiple instance-based video anomaly detection using deep temporal encoding-decoding", *Expert Systems with Applications*, Vol. 214, 2023, p.119079.
- [32]. S. Dubey, A. Boragule, M. Jeon, "3d resnet with ranking loss function for abnormal activity detection in videos", *Proceedings of the International Conference on Control, Automation and Information Sciences*, Chengdu, China, 23-26 October 2019, pp. 1-6.
- [33]. S. Dubey, A. Boragule, J. Gwak, M. Jeon, "Anomalous event recognition in videos based on joint learning of motion and appearance with multiple ranking measures", *Applied Sciences*, Vol. 11, No. 3, 2021, p.1344.
- [34]. T. Pevný, "Loda: Lightweight on-line detector of anomalies", *Machine Learning*, Vol. 102, 2016, pp. 275-304.

- [35]. M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, L. S. Davis, "Learning temporal regularity in video sequences", Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27-30 June 2016, pp. 733-742.
- [36]. G. Stein, U. Seljak, B. Dai, "Unsupervised in-distribution anomaly detection of new physics through conditional density estimation", arXiv:2012.11638, 2020.
- [37]. D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, "Learning spatiotemporal features with 3d convolutional networks", Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7-13 December 2015, pp. 4489-4497.
- [38]. J. Carreira, A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset", IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21-26 July 2017, pp. 6299-6308.
- [39]. S. Andrews, I. Tsochantaridis, T. Hofmann, "Support vector machines for multiple-instance learning", MIT Press, 2002, pp. 577-584.
- [40]. C. Lu, J. Shi, J. Jia, "Abnormal event detection at 150 fps in matlab", Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1-8 December 2013, pp. 2720-2727.
- [41]. J. Zhang, L. Qing, J. Miao, "Temporal convolutional network with complementary inner bag loss for weakly supervised anomaly detection", Proceedings of the IEEE International Conference on Image Processing, Taipei, Taiwan, 22-25 September 2019, pp. 4030-4034.
- [42]. J. Zhong et al. "Graph Convolutional Label Noise Cleaner: Train a Plug-And-Play Action Classifier for Anomaly Detection", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019 pp. 1237-1246.

Development of a Control Strategy for the Hybrid Energy Storage Systems in Standalone Microgrid

Original Scientific Paper

Hocine Guentri

Abdelhafid Boussouf University centre of Mila,
Mila, Algeria
Department of Mechanic and Electromechanics,
Institute of Science and Technologies,
GE Laboratory, Saida
Mila city, Algeria
hguentri2005@centre-univ-mila.dz

Abdeldjalil Dahbi

Unité de Recherche en Energies Renouvelables
en Milieu Saharien(URERMS), Centre de
Développement des Energies renouvelables(CDER),
01000, Adrar, Algeria
Laboratory of sustainable Development and
informatics, (L.D.D.I), University of Adrar
Adrar city, Algeria
dahbi_j@yahoo.fr

Tayeb Allaoui

Ibn Khaldoun university of TIARET,
Department of Electrical Engineering, Faculty of
Applied Science, L2GEGI Laboratory
Tiaret, Algeria
allaoui_tb@yahoo.fr

Salim Aoulmit

Abdelhafid Boussouf University centre of Mila,
Department of Mechanic and Electromechanics,
Institute of Science and Technologies,
Mila, Algeria
s.aoulmit@centre-univ-mila.dz

Ahmed Bouraiou

Unité de Recherche en Energies Renouvelables
en Milieu Saharien(URERMS), Centre de
Développement des Energies renouvelables(CDER),
01000, Adrar, Algeria
Laboratory of sustainable Development and
informatics, (L.D.D.I), University of Adrar
Adrar, Algeria
bouraiouahmed@gmail.com

Abstract – The intermediate energy storage system is very necessary for the standalone multi-source renewable energy system to increase stability, reliability of supply, and power quality. Among the most practical energy storage solutions is combining supercapacitors and chemical batteries. However, the major problem in this kind of application is the design of the power management, as well as the control scheme of hybrid energy storage systems. The focal purpose of this paper is to develop a novel approach to control DC bus voltage based on the reference power's frequency decomposition. This paper uses a storage system combined of batteries and supercapacitors. These later are integrated in the multi-source renewable energy system to supply an AC load. This technique uses the low-pass filters' properties to control the DC bus voltage by balancing the generated green power and the fluctuating load. The hybrid storage system regulates power fluctuations by absorbing surplus power and providing required power. The results show good performances of the proposed control scheme, such as low battery current charge/discharge rates, lower current stress level on batteries, voltage control improvements, which lead to increase the battery life.

Keywords: Photovoltaic, wind turbine generator, energy management, batteries, supercapacitor, hybrid energy storage

1. INTRODUCTION

Renewable energies become a very promote solution against environment problems and a suitable drawback for remote area. However, the intermitence criteria of these sources make the Standalone Multi-source Renewable Energy Generation Systems

(SMREGS) necessitate an intermediate energy storage system (ESS) to improve their stability, power quality, and supply reliability [1]. Chemical batteries, particularly lead-acid and lithium-ion batteries, are the most often utilized energy storage systems due to their high energy density and low cost relatively [1].

However, the primary drawbacks of these batteries are their limited lifespan and performance degradation in deep draining and overcharging cases. In order to enhance the capability and flexibility of the ESS, the combination of two or more energy storage technologies is a good solution in a hybrid configuration (HESS) [2].

Recently, researchers have shown increasing interest in investigating HESS and its applications. The most effective HESS configuration many researchers propose is the combination of chemical batteries with supercapacitors (SCs). Despite of SCs have a high power density, their energy density is relatively low [3]. They have a higher life cycle than batteries; they can also be charged and discharged quickly. Since each ESS has its drawbacks, the combining between the high density of the battery and the quick SC charging and discharging improves the overall reliability of the ESS.

However, the major problem with this kind of application is how to design the power management and control scheme of HESSs consisting of batteries and SCs [4]. Different control and management strategies have been proposed in the literature to stabilize the DC bus voltage and maintain a stable battery current during the transient time. Battery activities should be monitored and managed to operate safely continuously [5]. The most common techniques depend on a low-pass filter's frequency decomposition of the reference value, as proposed in [6, 7], or a discrete wavelet transform, as presented in [8, 9]. Other proposed model of predictive control-based methods, such as artificial neural networks [10], fuzzy supervisory control [11], and fuzzy adaptive control [12]. Authors in [13] used a genetic algorithm-based fuzzy logic controller to optimize the HESS. In [14], the adaptive power management of a standalone hybrid renewable energy system has been presented. In [15], the authors design an efficient energy management structure to improve the control of the HESS, which consists of batteries and supercapacitors. Also, in [16], the authors propose and develop the idea of using a community supercapacitor in an islanded DC multiple nano-grids system. In addition, in [17] the authors used only a PV source to provide electrical energy to a DC load, where the control strategy was based on the DC bus voltage regulation. However, in the current paper a hybrid system consisted of two sources (PV and wind) have been used, to supply an AC load via an inverter. Whereas, the control strategy is bases on the power regulation. The primary goal of this study is to develop a novel DC bus voltage management method based on the frequency decomposition of the reference power and a low-pass filter for the production of renewable energy from multiple sources. These sources supply an AC load with the help of a hybrid storage system composed of batteries and supercapacitors to improve the ESS and the stability of the DC bus voltage, power quality, and supply reliability.

The paper is organized as follows: Section 2 describes the system and the modeling of the different components. Section 3 presents the proposed energy man-

agement and control strategy of the HESS. Simulation results will be presented and discussed in Section 4. Finally, Section 5 summarizes this work's conclusions.

2. SYSTEM DESCRIPTION AND MODELING

The proposed renewable energy power system includes a wind turbine, a photovoltaic panel, and an ESS combining lithium batteries with an SC. As shown in Figure 1, all components are connected to a DC bus via a power converter. The whole system is connected to the AC load using a DC-AC converter.

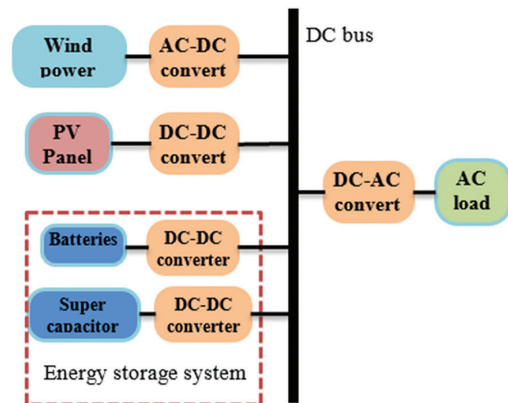


Fig. 1. Hybrid power system model

2.1. MODELING OF THE WIND ENERGY SYSTEM

The Wind Energy System is based on a permanent magnet synchronous generator (PMSG) coupled to the DC bus through a controlled AC-DC converter, it operates on variable speed mode with the pitch angle control as illustrated in Fig. 2 [18].

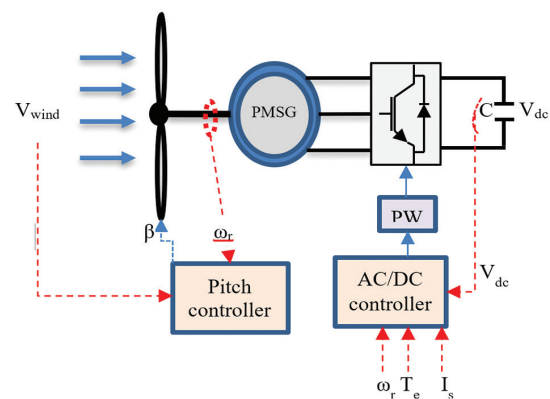


Fig. 2. Wind energy system configuration

2.1.1. Wind turbine model

The mechanical power produced by the wind turbine may be expressed as [19]:

$$P_m = \frac{1}{2} \rho \pi R^2 V_w^3 C_p(\lambda, \beta) \quad (1)$$

where P_m is the captured wind power (W), ρ is the air density (Kg/m³), R is the blade radius (m), V_w is the wind

speed (m/s), and C_p is the power coefficient. Based on the wind turbine characteristics, the value of C_p depends on the tip speed ratio (λ) and blade pitch angle (β) as:

$$C_p(\lambda, \beta) = c_1 \left(\frac{c_2}{\lambda_i} - c_3\beta - c_4 \right) e^{\frac{c_5}{\lambda_i}} + c_6\lambda \quad (2)$$

$$\frac{1}{\lambda_i} = \frac{1}{\lambda - 0.08\beta} - \frac{0.035}{\beta^3 + 1} \quad (3)$$

$$\lambda = \frac{\omega_r R}{V_w} \quad (4)$$

where c_1 to c_6 as given in Table 1, they denote the characteristic coefficients of the wind turbine.

Table 1. Wind turbine coefficients characteristics

c_1	c_2	c_3	c_4	c_5	c_6
0.5176	116	0.4	5	21	0.0068

2.1.2. The permanent magnet synchronous generator (PMSG) modeling

The stator voltage of the three-phase PMSG using the two-phase orthogonal rotating dq reference frame based on Park transformation is given by: [20, 21]:

$$\begin{cases} V_{ds} = I_{ds}R_s + L_{ds} \frac{dI_{ds}}{dt} - \omega_e L_{qs} I_{qs} \\ V_{qs} = I_{qs}R_s + L_{qs} \frac{dI_{qs}}{dt} + \omega_e (L_{ds} I_{ds} + \lambda_{pm}) \end{cases} \quad (5)$$

$$T_{em} = \frac{3Np}{4} (I_{qs}\lambda_{pm} - I_{ds}I_{qs}(L_{ds} - L_{qs})) \quad (6)$$

where L_{ds}, L_{qs} are the inductances in d and q axis, I_{ds}, I_{qs} , V_{ds}, V_{qs} represent respectively currents and voltages in d and q axis, R_s is the stator resistance, ω_e denotes the electric angular frequency, λ_{pm} is the permanent magnet flux, p is the number of pole pairs.

The PMSG electrical scheme in dq reference frame is shown in Fig. 3.

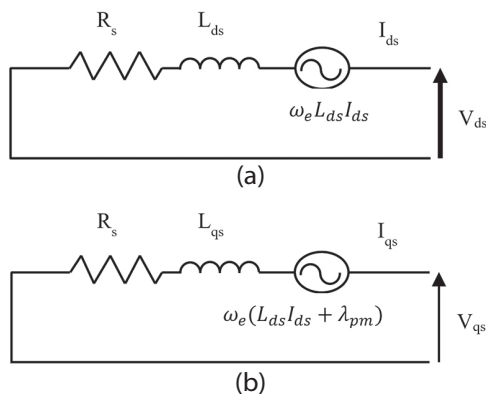


Fig. 3. PMSG (a) d-axis (b) q-axis circuit diagram

2.2. MODELING OF THE PHOTOVOLTAIC POWER GENERATOR

The photovoltaic (PV) system consists of a PV panel and a DC-DC boost converter controlled using the

perturb and observe Maximum Power Point Tracking (MPPT) technique as shown in Fig. 4. The DC-DC converter's function is used to modify the impedance to provide the optimum energy from the PV panel [22].

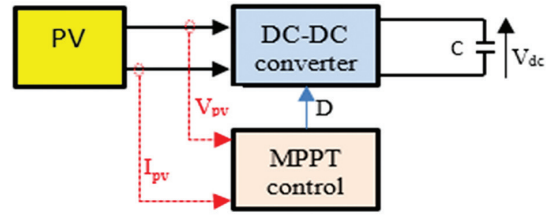


Fig. 4. Scheme of the photovoltaic system

Fig. 5 shows the commonly used model for a solar cell. It includes a photocurrent, a diode, a series resistor representing internal resistance to the passage of the current, and a parallel resistor representing the leakage current [23].

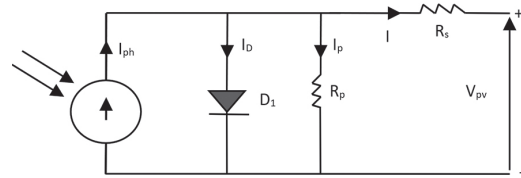


Fig. 5. PV cell equivalent circuit model

The equation of voltage-current characteristics of a solar cell's can be expressed by:

$$I = I_{ph} - I_s \left[\exp \left(\frac{q(V + IR_s)}{kT_c A} \right) - 1 \right] - \left(\frac{V + IR_s}{R_p} \right) \quad (7)$$

where I_{ph} is a photocurrent, I_s is the cell saturation of dark current, q is the electron charge ($q=1.6.10^{-19}C$), the Boltzman's constant ($k=1.3.10^{-23} J/K$), T_c is the cell's working temperature (Kalvin), and A is the ideality factor.

Equation (8) define the saturation current I_s as following:

$$I_s = I_p \left(\frac{T_c}{T_{ref}} \right)^3 \exp \left(qG_r \left(\frac{1}{T_{ref}} - \frac{1}{T_c} \right) \cdot \frac{1}{kA} \right) \quad (8)$$

where T_{ref} is the cell's reference temperature, and G_r is the solar irradiance.

The PV array comprises many PV modules that are electrically coupled in both parallel and serial circuits in order to produce the necessary current and voltage. With N_p, N_{sare} respectively number of parallels and series modules, the output or load current is calculated as [24]:

$$I = N_p I_{ph} - N_p I_s \left[\exp \left(q \left(\frac{V}{N_s} + \frac{IR_s}{N_p} \right) \frac{1}{kT_c A} \right) - 1 - \left(\frac{N_p V}{N_s} + IR_s \right) \frac{1}{R_p} \right] \quad (9)$$

2.3. MODELING OF THE BATTERIES

Battery modeling involves two steps: (1) the choice of a suitable model structure, (2) the determination of the model's parameter values. The RC ladder circuit shown in Fig. 6 is adopted in this research, because of

its excellent modeling accuracy and low computational complexity [25].

It is assumed that all cells are identical and the battery performance is unaffected by the packaging components. The connection of the cells (in series and parallel) does not affect the parameter values of the pack model. The battery pack connected in n-series has a fixed capacity. The battery pack model's open-circuit voltage (OCV) and R_s are computed as n times of the cell's R_s and OCV [6].

$$V_{bat} = OCV + \left(R_s + \sum_{i=1}^n Z_i \right) I_t \quad (10)$$

$$Z_i = R_i + \frac{1}{jC_i\omega} \quad (11)$$

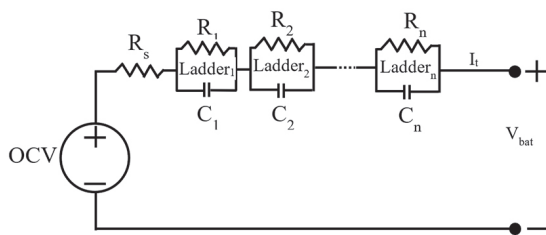


Fig. 6. The second RC ladder model structure

2.4. MODELING OF THE SUPERCAPACITORS

Various models of SC's have been reported in the literature based on their dynamic behaviour. The elementary series R-C is used in this work because of its simplicity and ability to replicate the actual SC's dynamic behaviour accurately. In [26], it is shown that SCs exhibit non-integer (fractional) behaviour and have a fractional impedance which is composed of a series resistor R_s and a fractional-order capacitor C_λ , as it is shown in Fig. 7.

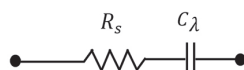


Fig. 7. Supercapacitors fractional series R-C model

The R-C series SC model's total equivalent impedance can be described as:

$$Z(s) = R_s + \frac{1}{C_\lambda s^\lambda} \quad (12)$$

Where C_λ is the pseudo-capacitance with order λ ($0 \leq \lambda \leq 1$).

3. ENERGY MANAGEMENT AND CONTROL STRATEGIES

The control strategy aims to accomplish the power flow of the HESS to reach the following main goals: (1) decreasing of dynamic battery stress levels; (2) maintaining a stable DC voltage; (3) avoiding deep discharge of the battery; and (4) improving the system's overall efficiency.

The HESS shown in Fig. 8 presents the combination of batteries with supercapacitors. Each storage device is coupled to the DC bus via a bidirectional buck-boost converter. HESS is used to preserve the constant DC bus voltage (V_{dc}) because of the imbalance between demand and multisource generation. If the multisource generation is lower than demand, the DC bus voltage V_{dc} decreases from its reference value. HESS would then discharge to satisfy the excess demand. Likewise, if the multisource generation exceeds the demand, the DC bus voltage increases from its reference value. As a result, HESS will charge to consume the extra power.

The block diagram of the control approach is shown in Fig. 9. The control allows batteries to support slow transients in the long term, whereas the SC supports fast transients in a short time [9].

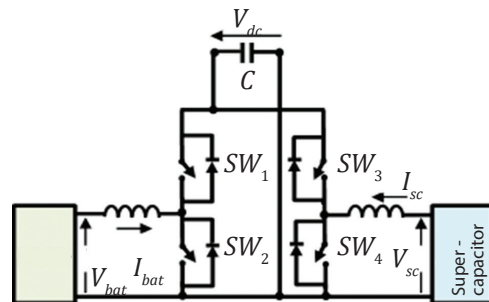


Fig. 8. HESS system

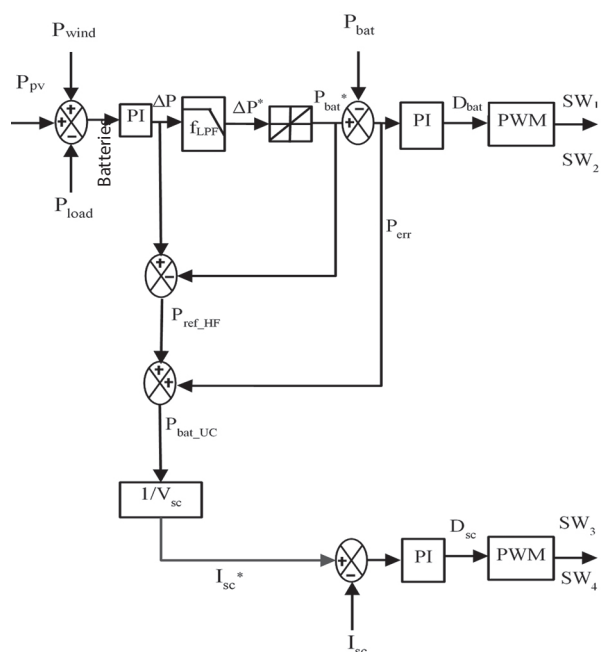


Fig. 9. The block diagram of the control strategy

The power deficit between the load demand and the generation from several sources, which is the total power that must be delivered by the HESS, it is given by:

$$\Delta P = P_{pv} + P_{wind} - P_{load} = P_{bat} + P_{sc} \quad (13)$$

Therefore, the ΔP value includes the high-frequency element, which the SC provides. However, batteries provide the low-frequency components.

In this work, a low pass filter has been proposed to determine the low-frequency element, which represents the deficit reference power, as given:

$$\Delta P^* = f_{LPF}(\Delta P) \quad (14)$$

Where, f_{LPF} is the low-pass filter function. The charge limiter is used to deliver batteries' reference power, which is provided by [27]:

$$P_{bat}^* = f_{LPF}(\Delta P^*) \quad (15)$$

The real value (P_{bat}) is compared with this reference. The PI controller receives the error to generate the duty cycle D_{bat} to adjust the PWM control signal for battery converter switches (SW1 and SW2), as seen in Fig. 9.

Furthermore, the high frequency component (P_{ref_HF}) is deduced as following.

$$P_{ref_HF} = \Delta P - P_{bat}^* \quad (16)$$

Therefore, the uncompensated battery power is:

$$P_{bat_UC} = P_{ref_HF} + P_{err} \quad (17)$$

The SC will supply the uncompensated battery power. As a result, the SC reference current will be:

$$I_{sc}^* = \frac{P_{bat_UC}}{V_{sc}} \quad (18)$$

A similar control strategy is applied to SC. The PI controller will produce the duty cycle D_{sc} to generate the PWM signal for the SW3 and SW4 supercapacitor converter switches, as seen in Fig. 9.

4. SIMULATION AND RESULTS

The complete model of the system described in Figure 1 has been simulated in a MATLAB/Simulink environment under various operation conditions to evaluate the performance of the proposed control and management strategy.

Table 2 (Appendix) lists the model parameters utilized in this simulation investigation. Where $K_{p,v}$, $K_{p,bat}$, $K_{p,sc}$ are the proportional gains for the DC bus voltage, batteries, and supercapacitor PI controllers, respectively, and $K_{i,v}$, $K_{i,bat}$, $K_{i,sc}$ are the integral gains of the DC bus voltage, batteries and supercapacitor PI controllers, respectively.

The batteries and supercapacitor are assumed to be initially charged. The simulations are performed under different operating conditions, including variable irradiance, variable wind speed, and variable load.

Case: 1 Variable solar irradiance, constant wind speed and constant load

The solar irradiance profile employed in this simulation scenario is shown in Fig. 10, whereas the wind speed and load are fixed.

The performance of the proposed management and control strategy under variable solar irradiation is shown in Fig. 11 and 12. The load requests always 19 kW, as seen from (0 s to 1 s): $P_{pv} = 10$ kW, the wind power provides about

4.9 kW, while the battery supplies the remaining power. Similarly, the battery power follows the PV power variations. The battery is in discharge mode, providing insufficient power until reaching the load power request (19 kW). Alternatively, when there is an excess of power, the battery is in charging mode and stores the extra energy.

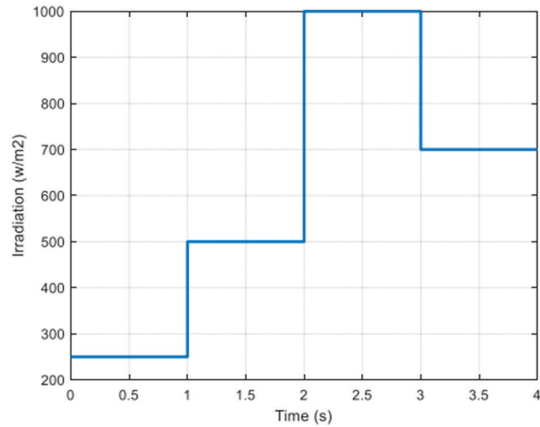


Fig. 10. Solar irradiance

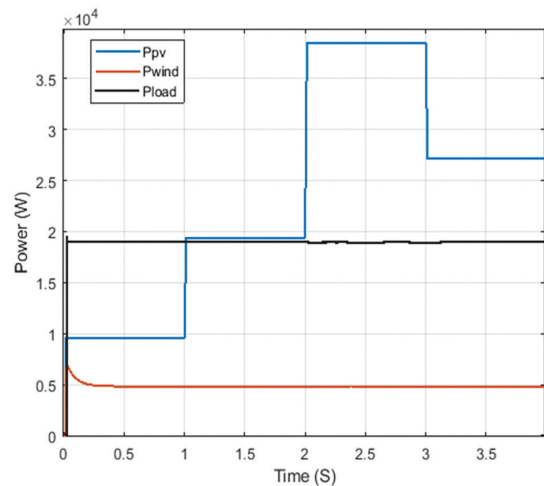


Fig. 11. Simulation of P_{pv} , P_{wind} and P_{load} with variable solar irradiance

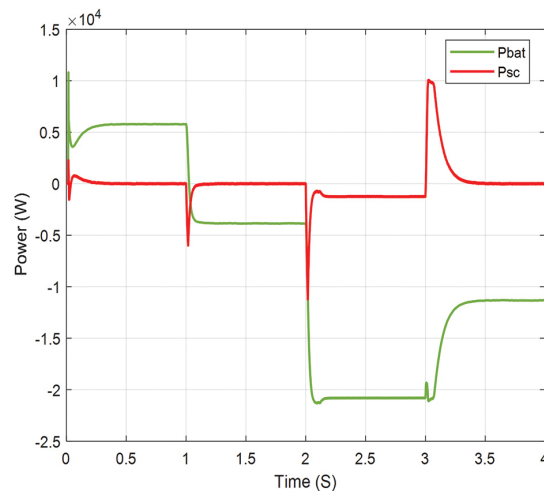


Fig. 12. Simulation of P_{pv} , P_{wind} and P_{load} with variable solar irradiance

Case: 2 Variable wind speed with a constant solar irradiation and constant load.

Fig. 13 illustrates the wind speed profile used in this simulation scenario. In this case, the solar irradiance and loads are kept constant.

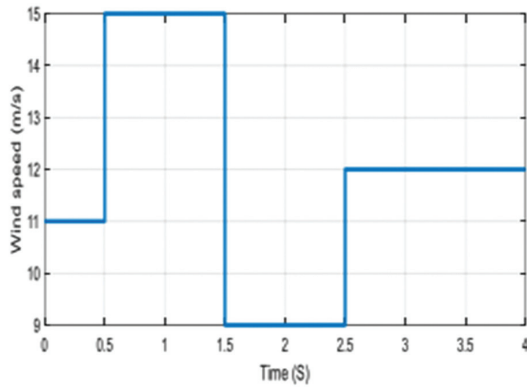


Fig. 13. Wind speed profile

Fig. 14 and 15 illustrate the proposed management and control strategy's performances under variable wind speeds.

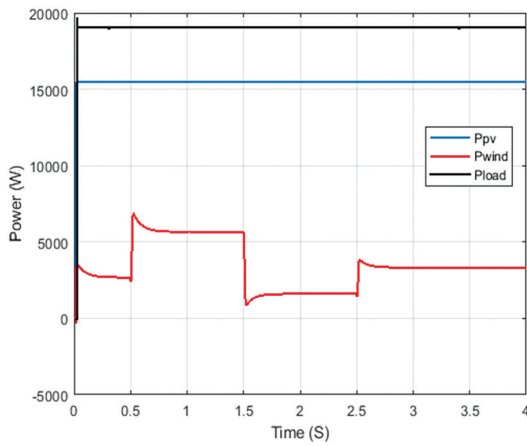


Fig. 14. Simulation of P_{pv} , P_{wind} and P_{load} with variable wind speed.

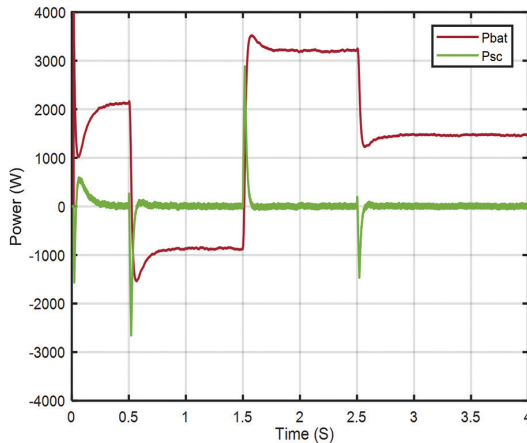


Fig. 15. Simulation of P_{bat} and P_{sc} with variable wind speed

The wind power supplies about 3 kW from (0 s to 1s), the PV panel provides 15 kW, and the battery delivers the rest of the power needed by the load. Likewise, the battery power follows the wind power variations. When there is insufficient electricity, the battery discharges and provides the required energy. Alternatively, the battery is in charging mode and stores any excess power from several sources when there is a surplus.

Case: 3 Variable load with a constant solar irradiation and constant wind speed.

Fig. 16 shows the load profile considered in this simulation with a constant solar irradiance and wind speed.

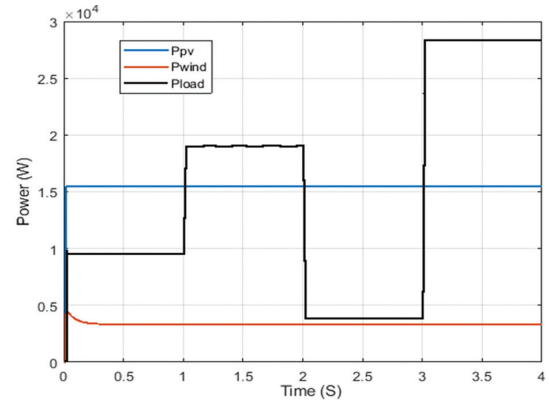


Fig. 16. Simulation of P_{pv} , P_{wind} and P_{load} under variable load.

The solar irradiance is set to 400 W/m², where the photovoltaic cell power is equal to P_{pv} =16 kW, and the wind speed is 14 m/s, where the wind power is P_{wind} =6 kW. Similarly, the battery power follows the load variations. When there is insufficient electricity, the battery discharges and supplies the needed power. Alternatively, the battery is in charging mode and stores extra multi-source generation when there is an excess of power, as shown in Fig. 16 and 17.

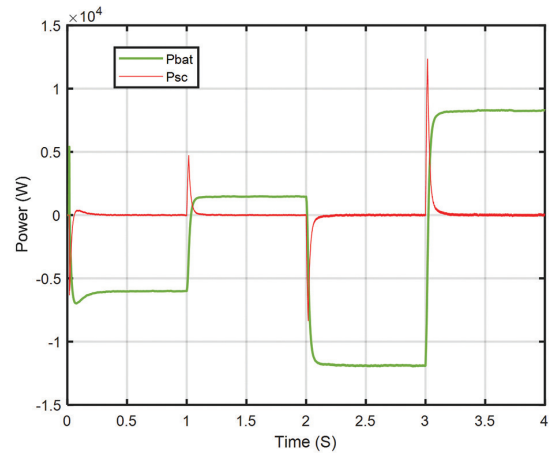


Fig. 17. Simulation of P_{bat} and P_{sc} under variable load

Figure 18 illustrates the SoC of the supercapacitor variation under variable load. As it can be seen, there is a similarity between this profile and the profile of

supercapacitor power variation, because the supercapacitor's SoC follows the latter's power variation.

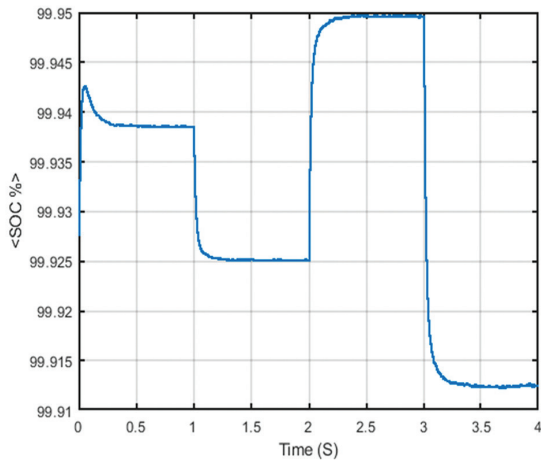


Fig. 18. Simulation of SoC_{sc} under variable load

This control strategy also shows its performance by keeping the DC bus voltage closer to its reference value $V_{dc}=400$ V, as shown in Fig. 18. At $(t=2$ s), the load decreases from 20 kW to 4 kW. It can be observed on the zoomed-in section of the waveform in Fig. 19 that the DC bus voltage remains stable, this is on one hand.

On the other hand, Fig. 20 shows that the DC bus voltage is unstable and follows the fluctuations of the load or the different sources.

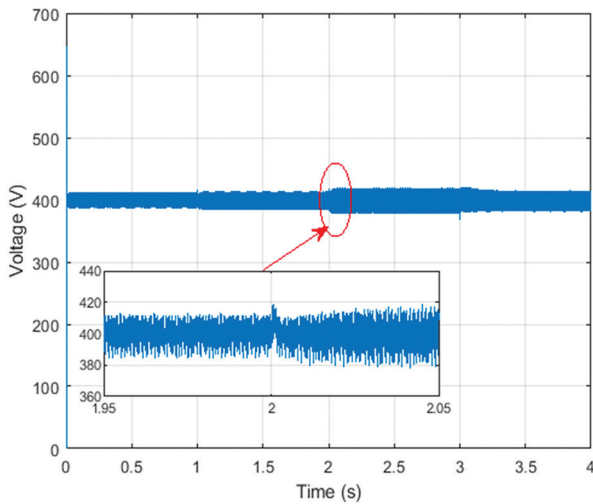


Fig. 19. Simulation of V_{dc} with power management, under variable load

Based on the three scenarios considered in this simulation study, the results of this study indicate that the batteries and supercapacitors can respond properly to the fluctuations of the electrical load demand. The SCs quickly provide or absorb peak powers in response to the load demands, to compensate the slow batteries' delay relatively. The current stress level on batteries is significantly reduced. The DC bus voltage is successfully maintained constant at 400 V, as required by the control and management strategy.

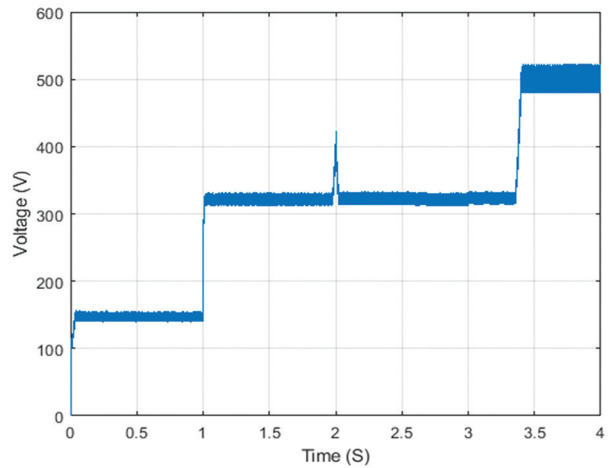


Fig. 20. Simulation of V_{dc} without power management, under variable load

5. CONCLUSION

The primary objective of the current study was to develop and evaluate a control strategy based on the regulation of the low-frequency power components of multisource generating systems and the load imbalance. Detailed modeling of the system's main components, such as photovoltaic and wind power, has been presented. Then, the proposed control strategy was detailed and tested. The results of this investigation, through the three simulation scenarios, prove the validity and goals of the control strategy: reducing of battery charge/discharge rates, improvement of voltage regulation, and the electric current causing lower stress levels on the battery, which increases the battery life. Future work can be focused on testing experimentally these microgrid performances under partial shading conditions.

6. APPENDIX

Table 2. System parameters used in the simulation study

Parameters	Values
PV array	
Open circuit voltage (V_{pv})	21 V
Short circuit current (V_{pv})	8 A
Number of Parallel strings (N_p)	20
Number of Series-connected modules per string (N_s)	16
Maximum Power (W)	120.7
Voltage at maximum power point V_{mp} (V)	17
Current at maximum power point I_{mp} (A)	7.1
Wind energy system	
Stator phase resistance R_s (Ω)	2.875
Armature inductance (H)	0.000835
Flux linkage (Wb)	0.175
Number of pole pairs	4
Electrical power generator (kW)	6

Battery specifications	
Type	Lithium-Ion
Ah capacity	45
Terminal voltage (V_{bat})	12
Number of batteries in series	30
SC specifications	
Terminal voltage (V_{sc})	50
Capacitance (C_{sc})	25 F
Number of SCs in series	4
Converters parameters	
L_{pv}	0.352 mH
C_{pv}	100 μ F
L_{bat}	0.3 mH
L_{sc}	0.355 mH
C_{dc}	7000 μ F
L_f	5 mH
C_f	10 μ F
PI controllers parameters	
K_{pv}	1.477
K_{pbat}	3077
K_{psc}	0.043
K_{iv}	0.65
K_{ibat}	0.45
K_{isc}	100
DC and AC grid parameters	
V_{dc}	400 V
V_{ac}	220 V
f	50 Hz

7. NOMENCLATURE

EES	electrical energy storage	R_s	series resistance (Ω)
$HEES$	hybrid electrical energy storage	R_p	parallel resistance (Ω)
SC	supercapacitor	I	solar cell current (A)
$PMSG$	permanent magnet synchronous generator	I_{ph}	Photocurrent (A)
P_m	captured wind power	I_s	cell saturation of dark current (A)
ρ	the air density (Kg/m ³)	I_D	The diode current (A)
R	the radius of the rotor blade (m)	V_{pv}	solar cell voltage (V)
V_w	wind speed(m/s)	q	the electron charge (C)
C_p	the power coefficient	k	the Boltzman's constant (J/K)
PMS	power management strategy	T_c	the cell's working temperature
PV	photovoltaic	A	the ideality factor
DC	Direct current	I_p	parallel current (A)
$MPPT$	maximum power point tracking	T_{ref}	the cell's reference temperature ($^{\circ}$ C)

PI	Proportional integrator	G_r	the solar insulation (W/m ²)
N_p	Parallels modules	V_{ESS}	The voltage of the electrical energy storage (V)
N_s	series modules	V_{dc}	DC bus voltage (V)
C	DC bus capacitor (F)	P_{wid}	The wind power generator
R	The load resistance (Ω)	L_{pv}	The DC-DC converter inductance of the PV (H)
D_b	the duty ratio of the control of the DC/DC converter of the battery	L_b	The DC-DC converter inductance of the batteries (H)
SC	supercapacitor	L_{sc}	The DC-DC converter inductance of the supercapacitor (H)
D_{sc}	the duty ratio of the control of the DC/DC converter of the SC	I_{load}	The load current (A)
$I_{dc,ref}$	The DC bus current reference (A)	$V_{dc,ref}$	The DC bus reference voltage (V)
$I_{sc,ref}$	The SC current reference (A)	K_{p-sc}	The proportional factor of the SC corrector
$I_{bat,ref}$	The batteries current reference (A)	K_{p-bat}	The proportional factor of the battery corrector
V_{sc}	The supercapacitor voltage (V)	K_{i-sc}	The integrator factor of the SC corrector
V_{bat}	The battery voltage (V)	K_{i-bat}	The integrator factor of the battery corrector
I_{sc}	The SC current (A)	K_{p-DC}	The proportional factor of the DC bus corrector
I_{bat}	The battery current (A)	K_{i-DC}	The integrator factor of the DC bus corrector
PID	Proportional integrator derivator	P_{sc}	The SC power
P_{bat}	batteries power	P_{pv}	PV power
P_{load}	load power		

8. REFERENCES:

- [1] K. C. Divya, J. Ostergaard, "Battery energy storage technology for power systems — An overview", Electric Power Systems Research, Vol. 79, 2014, pp. 511-520.
- [2] T. Rout, M. K. Maharana, A. Chowdhury, S. Samal, "A Comparative study of Stand-alone Photo-Voltaic System with Battery storage system and Battery Supercapacitor storage system", Proceedings of the 4th International Conference on Electrical Energy Systems, Chennai, India, 7-9 February 2018, pp. 77-81.
- [3] Y. Zhang, Z. Jiang, X. Yu, "Control Strategies for Battery / Supercapacitor Hybrid Energy Storage Systems", Proceedings of the IEEE Energy 2030 Conference, Atlanta, GA, USA, 17-18 November 2008, pp. 5-10.

- [4] A. Ferrara, C. Hametner, "Impact of Energy Management Strategies on Hydrogen Consumption and Start-Up/Shut-Down Cycles in Fuel Cell-Ultracapacitor-Battery Vehicles", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 6, 2022, pp. 5692-5703.
- [5] M. O. Qays, Y. Buswig, M. L. Hossain, A. Abu-Siada, "Recent progress and future trends on the state of charge estimation methods to improve battery-storage efficiency: A review", *CSEE Journal of Power and Energy Systems*, Vol. 8, No. 1, 2022.
- [6] S. A. Ghorashi, K. Abadi, S. I. Habibi, T. Khalili, A. Bidram, "A Model Predictive Control Strategy for Performance Improvement of Hybrid Energy Storage Systems in DC Microgrids", *IEEE Access*, Vol. 10, 2022, pp. 25400-25421.
- [7] S. A. Ghorashi, K. Abadi, A. Bidram, "A distributed rule-based power management strategy in a photovoltaic/hybrid energy storage based on an active compensation filtering technique", *IET Renewable Power Generation*, Vol. 15, No. 15, 2021, pp. 3688-3703.
- [8] M. Robayo, M. Mueller, S. Sharkh, M. Abusara, "Assessment of supercapacitor performance in a hybrid energy storage system with an EMS based on the discrete wavelet transform", *Journal of Energy Storage*, Vol. 57, 2023, p. 106200.
- [9] N. M. Ismail, M. K. Mishra, "A multi-objective control scheme of a voltage source converter with battery-supercapacitor energy storage system used for power quality improvement", *International Journal of Electrical Power & Energy Systems*, Vol. 142A, 2022, p. 108253.
- [10] S. Xie, X. Hu, S. Qi, K. Lang, "An artificial neural network-enhanced energy management strategy for plug-in hybrid electric vehicles", *Energy*, Vol. 163, 2018, pp. 837-848.
- [11] P. Mohan R., R. S. Sreelekshmi, M. G. Nair, "A Neural Network Based Hybrid Energy Storage System Connected to a DC Standalone PV System", *Proceedings of the International Virtual Conference on Power Engineering Computing and Control: Developments in Electric Vehicles and Energy Sector for Sustainable Future*, Chennai, India, 2022, pp. 1-5.
- [12] A. M. Abomazid, N. A. El-Taweel, H. E. Z. Farag, "Optimal Energy Management of Hydrogen Energy Facility Using Integrated Battery Energy Storage and Solar Photovoltaic Systems", *IEEE Transactions on Sustainable Energy*, Vol. 13, No. 3, 2022, pp. 1457-1468.
- [13] M. Faisal, M. A. Hannan, P. J. Ker, K. M. Muttaqi, "Genetic Algorithm based Fuzzy Logic Controller for Optimal Charging-Discharging of Energy Storage in Microgrid applications", *Proceedings of the IEEE Industry Applications Society Annual Meeting*, Detroit, MI, USA, 2022, pp. 1-9.
- [14] K. R. Naik, B. Rajpathak, A. Mitra, M. Kolhe, "Adaptive Energy Management Strategy for Optimal Power Flow Control of Hybrid DC Microgrid", *Proceedings of the 5th International Conference on Smart and Sustainable Technologies*, Split, Croatia, 2020, pp. 1-6.
- [15] A. Bharatee, P. K. Ray, A. Ghosh, "A Power Management Scheme for Grid-connected PV Integrated with Hybrid Energy Storage System", *Journal of Modern Power Systems and Clean Energy*, Vol. 10, No. 4, 2022, pp. 954-963, July 2022.
- [16] S. A. Ghorashi, K. Abadi, A. Bidram, "Effective utilization of grid-forming cloud hybrid energy storage systems in islanded clustered dc nano-grids for improving transient voltage quality and battery lifetime", *IET Renewable Power Generation*, Vol. 17, No. 8, 2023, pp. 1836-1856.
- [17] H. Guentri, T. Allaoui, M. Mekki, M. Denai, "Power management and control of a photovoltaic system with hybrid battery-supercapacitor energy storage based on heuristics methods", *Journal of Energy Storage*, Vol. 39, 2021, p. 102578.
- [18] B. Peng, X. Ma, X. Ma, C. Tian, "Fuzzy-based coordinated control and parameter correction strategy for speed controller of PMSG wind turbine in frequency response", *Energy Reports*, Vol. 9, No. 3, 2023, pp. 558-566.
- [19] K. Palanimuthu, G. Mayilsamy, S. R. Lee, S. Y. Jung, Y. H. Joo, "Comparative analysis of maximum power extraction and control methods between PMSG and PMVG-based wind turbine systems", *Energy Reports*, Vol. 143, 2022, p. 108475.
- [20] Y. Zhu, Z. Wang, X. Guo, Z. Wei, "An improved kinetic energy control strategy for power smooth-

- ing of PMSG-WECS based on low pass filter and fuzzy logic controller", *Electric Power Systems Research*, Vol. 214A, 2023, p. 108816.
- [21] Y. Belkhier, A. Achour, N. Ullah, R. N. Shaw, S. Chowdhury, K. Techato, "Energy-based fuzzy supervisory non integer control for performance improvement of PMSG-Based marine energy system under swell effect and parameter uncertainties", *Renewable Energy*, Vol. 186, 2022, pp. 457-468.
- [22] G. Liu, T. Li, J. Su, "Photovoltaic fault diagnosis method based on BP neural network and rule-based control chart", *Proceedings of the China International Conference on Electricity Distribution*, Changsha, China, 2022, pp. 579-583.
- [23] F. Bradaschia, M. C. Cavalcanti, A. J. do Nascimento, E. A. da Silva, G. M. de Souza Azevedo, "Parameter Identification for PV Modules Based on an Environment-Dependent Double-Diode Model", *IEEE Journal of Photovoltaics*, Vol. 9, No. 5, 2019, pp. 1388-1397.
- [24] B. Guddanti, J. R. Orrego, R. Roychowdhury, M. S. Illindala, "Sensitivity Analysis Based Identification of Key Parameters in the Dynamic Model of a Utility-Scale Solar PV Plant", *IEEE Transactions on Power Systems*, Vol. 37, No. 2, 2022, pp. 1340-1350.
- [25] P. V. H. Seger, P. X. Thivel, D. Riu, "A second life Li-ion battery ageing model with uncertainties: From cell to pack analysis", *Journal of Power Sources*, Vol. 541, 2022, p. 231663.
- [26] K. Milić, S. M. Cvetičanin, N. Vukajlović, D. Milicević, M. Subotin, D. Pejić, "Fractional modeling and analytic solution of supercapacitor response", *Proceedings of the IEEE PES Innovative Smart Grid Technologies Conference Europe*, Novi Sad, Serbia, 2022, pp. 1-5.
- [27] M. Keerthana, P. S. Manoharan, A. Ravi, "Design of Control Strategy for Battery-Supercapacitor Hybrid Storage System", *Proceedings of the 7th International Conference on Communication and Electronics Systems*, Coimbatore, India, 2022, pp. 798-803.

Performance of Synchronous Reluctance Generators with Series and Shunt Stator Connections

Original Scientific Paper

Pauline Ijeoma Obe

University of Nigeria Nsukka
Department of Industrial Technical Education,
Faculty of Vocational Technical Education
Nsukka 410001, Nigeria
pauline.obe@unn.edu.ng

Lilian Livutse Amuhaya

Botswana International University of Science and
Technology, Department of Electrical, Electronic
and Telecommunications Engineering,
Plot 10071 Boseja, Palapye, Botswana
amuhaya@biust.ac.bw

Emeka Simon Obe

Botswana International University of Science and
Technology
Department of Electrical, Electronic and
Telecommunications Engineering,
Plot 10071 Boseja, Palapye, Botswana
obee@biust.ac.bw

Adamu Murtala Zungeru

Botswana International University of Science and
Technology, Department of Electrical, Electronic
and Telecommunications Engineering,
Plot 10071 Boseja, Palapye, Botswana
zungerum@biust.ac.bw

Abstract – This paper reports the performance of series- and shunt-connected self-excited reluctance generators (SERG). In addition to the two stator connections, an analysis was carried out on rotor configurations (with and without a cage) a combination resulting in four different generator topologies. The loss of load and transient characteristics of each generator configuration were studied for a combination of pure resistive and R-L loads. It is shown that for the same machine size, speed and exciting capacitor value, the generator with a cage preserves a better wave shape following a transient disturbance than the cageless machine. At unity power factor, shunt generator with cage can deliver 0.691 pu output power, at 1.97% regulation; its series counterpart only delivers 0.589 pu at 2.05%. The study demonstrates that while shunt generators have better regulation and supports higher loads at different power factors, series generators show a superior performance in terms of damping out transients.

Keywords: synchronous reluctance generator, transient performance, voltage regulation, loss of load performance, series connection, shunt connection, exciting capacitor

1. INTRODUCTION

Generating operation of the synchronous reluctance machine was first reported by Abdel-Kader in 1985 [1]. The machine has not found much industrial application in power generation because it produces an output power that is less than that of an induction generator of comparable size and much less than that of field-excited generators. A major advantage of this generator type is that for loads within its carrying capacity, the frequency of the voltage developed has a well-defined relationship with the number of poles and operating speed of the generator. Induction generators are more prominent than reluctance generators because for the same machine size, speed and terminal excitation capacitance, it produces higher terminal voltage and higher current at dropout-voltage [2] and hence higher maximum power than the reluctance generator. This

is because it has a relatively higher magnetizing reactance. Unfortunately, frequency of induction generator voltages depends on load power factor and capacitor values and preservation of voltage wave shapes for all load conditions requires additional control circuitry which increases its cost. Reluctance generators are constant frequency devices. This has excited interest in research on SERG. In [3], the machine characteristics was reported; the steady state analysis has been performed [4-6] developing equations for the d-axis reactance and capacitance required or excitation in terms of load angle.

The stability and its response to small perturbations [7] and dynamic effects were presented in [8]. Boldea, *et al* [9] used a high reactance ratio axially-laminated anisotropic (ALA) machine and recommended its suitability as a source of controlled dc power. A series-con-

nected generator was reported in stand-alone mode by Ben-hail and Rabin [10]. The impact of cross-saturation was presented by Vagati, *et al* [11] for a transverse-laminated (TLA) rotor configuration.

The performance of SERG with- and without rotor cage has been reported in [12]. A recent paper by Ibrahim and Pillay [13] that dwelt deeply on SERG excitation in stand-alone mode developed a hysteresis current characteristic from which the voltage build-up and loss of excitation due to short-circuits for shunt-excited generators can be predicted. The paper produced a hysteresis loop of the machine and gave a clearer insight into the collapse of residual magnetism. The saturation curves needed for simulation in SERG were studied in detail by Guha and Kar [14] and Hoffer, *et al* [15] by considering several models, yielding the stability limits of the generator. Wang and Bianchi [5], inserted magnets into the ALA barriers and reported improved performance than conventional SERGs. To initiate self-excitation, it was emphasized that even in the presence of sufficient residual magnetism a closed loop needs to exist between the windings and a capacitor [16, 17]. Even when a suitable capacitance is connected, the residual magnetism has to be sufficient not just for the initiation of self-excitation but to sustain it when there are seemingly tolerable load changes [18].

The paper by Sekdy, *et al* [19] reported a voltage and frequency control of SERG using adjustable (switching) capacitors connected in series and shunt to load in an existing and already connected conventional shunt excitation fixed capacitor resulting in an improvement on voltage regulation for optimal load capacitor values. The paper by Hong, *et al*, [20] described a unique cooling method, loss calculation and parameter sensitivity analysis for a high speed SERG assisted by permanent magnets for use in aviation applications.

In [21], the electromagnetic design and analysis of a five phase reluctance generator with permanent magnet excitation with rectifier load was reported as having a higher power density than the conventional three phase machine. Grid applications of related generators have been explored in [22, 23].

From the foregoing, SERG with cage and with shunt capacitor connection appears to have received more attention [1-9, 12-19] than the series-excited machine cageless types [10, 11]. No attempt to the best of knowledge of the authors has been made to study both types and place their performance side-by-side and this forms the basis of this paper.

Section 2 deals with the general description of the stator connections and rotor configuration while section 3 details the mathematical modeling of the generator topologies. In section 4, the machine self-excitation process is presented and section 5 discusses the transient performance of the generators following the addition and removal of loads. Section 6 presents their voltage regulation and loadability characteristics while the study is concluded in section 7.

2. THE MACHINE UNDER STUDY

The connection diagram for both series and shunt-excited machines for one phase are shown in Fig. 1.

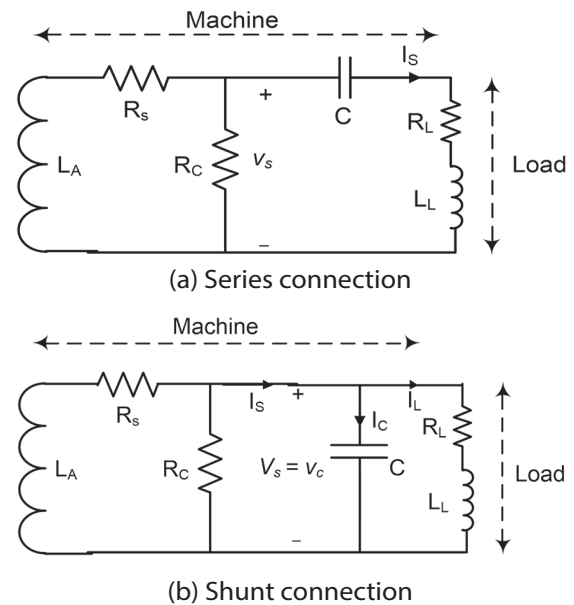


Fig. 1. Stator connections including core loss

The rotor shape is transverse-laminated (TLA) type. The stator winding is uniformly distributed around 36 slots of the machine and short-pitched 7/9 for all the generators to yield a more sinusoidal waveform. The stator winding layout over one pole is shown in Fig. 2.

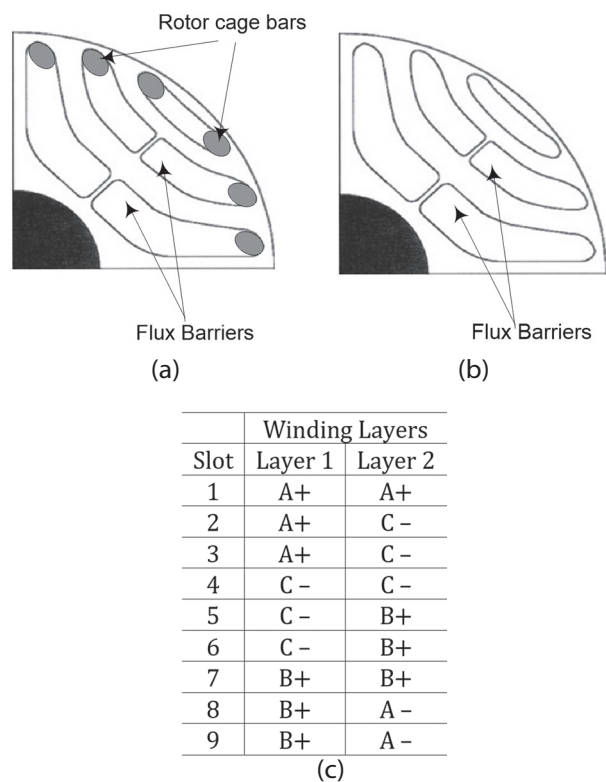


Fig. 2. Machine under study; (a) rotor with cage, (b) rotor without cage, (c) Stator winding layout over one pole-pitch

3. 3. THE SYSTEM MODEL

3.1. MACHINE MODEL

The dynamic voltage equations of a SERG with distributed polyphase stator winding and with a short-circuited rotor cage winding is given by [12]:

$$\begin{aligned} v_{qs} &= -R_s i_{qs} + \omega_r \lambda_{ds} + \frac{d}{dt} \lambda_{qs} \\ v_{ds} &= -R_s i_{ds} - \omega_r \lambda_{qs} + \frac{d}{dt} \lambda_{ds} \\ v_{qr}' &= R_{qr}' i_{qr}' + \frac{d}{dt} \lambda_{qr}' \\ v_{dr}' &= R_{dr}' i_{dr}' + \frac{d}{dt} \lambda_{dr}' \end{aligned} \quad (1)$$

where, i_{qs} , i_{ds} , i_{qr}' and i_{dr}' are the stator and rotor q - and d -axis currents respectively, λ_{ds} , λ_{qs} , λ_{dr}' and λ_{qr}' are the stator and rotor flux linkages, ω_r is the rotor speed and R_{qr}' and R_{dr}' are the rotor winding resistances in the equivalent q - and d -axis respectively.

The resistance R_s appearing in the stator windings are values already accounted for by core loss resistance R_c according to $R_s = (R_a \times R_c) / (R_a + R_c)$ where R_c is the core loss and R_a is the stator winding resistance. The flux linkages are defined as:

$$\begin{aligned} \lambda_{qs} &= (L_{ls} + L_{mq}) i_{qs} + L_{mq} i_{qr}' \\ \lambda_{ds} &= (L_{ls} + L_{md}) i_{ds} + L_{md} i_{dr}' \\ \lambda_{qr}' &= (L_{lqr}' + L_{mq}) i_{qr}' - L_{mq} i_{qs} \\ \lambda_{dr}' &= (L_{ldr}' + L_{md}) i_{dr}' - L_{md} i_{ds} \end{aligned} \quad (2)$$

Here, L_{ls} is the leakage inductance in the stator while L_{lqr}' and L_{ldr}' are the rotor leakage inductances in q - and d - axis respectively, L_{mq} and L_{md} are the magnetizing inductances in the q - and d - axis respectively. The mechanical equation of the rotor is given by:

$$\begin{aligned} \omega_r &= \frac{p_r}{2J} \int (T_e - T_L) dt \\ T_e &= \frac{3}{4} p_r (i_{qs} (i_{ds} + i_{dr}') L_{md} - i_{ds} (i_{qs} + i_{qr}') L_{mq}) - T_L \end{aligned} \quad (3)$$

Equations (1) – (3) are also valid for a cageless machine if all the rotor quantities i_{qr}' , i_{dr}' , λ_{dr}' and λ_{qr}' are set to zero.

3.2 CAPACITOR AND LOAD MODELS

The loads on each of the generators will be loads of lagging power factor (R-L series type). Considerations were not made about leading power factors because such loads are not very common in practice. For a changing R_L , the corresponding inductance L_L can be calculated for a specified power factor load using:

$$L_L = \frac{R_L}{\omega_r} \times \sqrt{\left(\frac{1}{\cos\phi}\right)^2 - 1} \quad (4)$$

3.2.1. Capacitor and load model for the series generator

If the capacitor is connected to the generator as shown in Fig. 1(a), it will be easy to write both the ca-

pacitor and load voltage equation as seen by the machine terminal voltage, v_s as:

$$v_s = i_s R_A + \frac{1}{C} \int i_s dt + L_L \frac{di_s}{dt} \quad (5)$$

It is obvious that the stator winding current i_s flows directly through the load. The resistance R_A appearing in (5) represents the stator winding resistance R_s and the load resistance R_L . Noting that R_L , L_L and C are constants and equal in each of the three phases, a transformation of (4) to the d - q reference frame fixed to the rotor yields:

$$\begin{aligned} v_{qs} &= R_A i_{qs} + \frac{i_{ds} - C \frac{dv_{dc}}{dt}}{\omega_r C} + L_L \frac{di_{qs}}{dt} + \omega_r L_L i_{qs} \\ v_{ds} &= R_A i_{ds} + \frac{i_{qs} + C \frac{dv_{dc}}{dt}}{\omega_r C} + L_L \frac{di_{ds}}{dt} - \omega_r L_L i_{ds} \end{aligned} \quad (6)$$

From (6), the volage across the load in Fig. 1(a) may be calculated.

3.2.2. Capacitor and load model for the shunt generator

For the connection of Fig. 1b, the voltage across the capacitor is also the voltage across the load. The equation of this voltage in stator variable is:

$$v_s = v_c = \frac{1}{C} \int (i_s - i_L) dt \quad (7)$$

When (7) is transformed to the rotor reference frame, we have:

$$\begin{aligned} v_{qs} &= \frac{1}{\omega_r C} \left(i_{dL} - i_{ds} + C \frac{dv_{ds}}{dt} \right) \\ v_{ds} &= \frac{1}{\omega_r C} \left(i_{qL} - i_{qs} - C \frac{dv_{qs}}{dt} \right) \end{aligned} \quad (8)$$

Equation (8) shows that the exciting current charging the capacitor and maintaining the stator flux is supplied by the difference between the winding currents and the load currents, hence only a fraction of the winding current is available to the load. The load currents are modeled as supplying a general R-L load and are given as:

$$\begin{aligned} i_{qL} &= \frac{1}{\omega_r L_L} \left(R_L i_{dL} - v_{ds} - L_L \frac{di_{dL}}{dt} \right) \\ i_{dL} &= \frac{1}{\omega_r L_L} \left(R_L i_{qL} - v_{qs} + L_L \frac{di_{qL}}{dt} \right) \end{aligned} \quad (9)$$

3.4. INDUCTANCE SATURATION AND CORE-LOSS MODEL

The d - and q - axis inductances calculated by finite element in the presence or absence of a cage gave similar saturation and core-loss data for both rotor structures on the ground that a constant speed operation is assumed. These similarities create a uniform basis for the comparative studies on the performance of the four generators studied here. Equation (10) is a third order polynomial fit showing the dependence of direct- and quadrature-axis inductances and core loss on the magnetizing flux linkage of the machine, obtained with finite element software FEMAG DC®.

$$L_d = -3.3\lambda_m^3 + 1.5\lambda_m^2 - 0.26\lambda_m + 0.231 H$$

$$L_q = -2.2\lambda_m^3 + 2.5\lambda_m^2 - \lambda_m + 0.17 H \quad (10)$$

$$R_c = -18800\lambda_m^3 + 13300\lambda_m^2 - 1980\lambda_m + 669 \Omega$$

where $\lambda_m = \sqrt{\lambda_{qs}^2 + \lambda_{ds}^2}$

4. VOLTAGE BUILD-UP

A common requirement for self-excitation of SERG is the existence of appropriate values of magnetic remanence, capacitor, speed and saturation. The constraint in the choice of these initial conditions is to ensure that the thermal limits and current carrying capability of the stator windings are not exceeded and that the generated voltage and frequency is nominally a universal value. Theoretically a large range of voltages and frequency can be generated whenever self-excitation occurs. Nearly all the papers cited emphasized these and, in this paper, a common set of values were chosen for the four configurations to ensure a fair comparison. To achieve this, a remanent flux of 0.01wb is assumed in the core, a rotor speed of $\omega_r = 377$ rads/sec and a constant excitation capacitor of 65 μ F were used.

The solution of the above equations is accomplished with Matlab Simulink[®]. In addition to (4) and (10) which must be used during the simulation of all the self-excited generator types, the following combination of equations and modifications were used for the simulation of the four generator topologies as shown in Table 1.

Table 1. Generator topologies under study

Assigned Code	Generator Topology	Equations for Simulation
RG1	Shunt stator connection with rotor cage	The whole of (1), (2), (8) and (9).
RG2	Shunt stator connection with cageless rotor	The first two equations of (1) and first two of (2) with i_{qr}' and i_{dr}' set to zero, and the whole of (8) and (9).
RG3	Series stator connection with rotor cage.	The whole of (1), (2) and (6).
RG4	Series stator connection with cageless rotor	The first two equations of (1) and first two of (2) with i_{qr}' and i_{dr}' set to zero, and the whole of (6).

When these sets of equations are separately solved under no-load conditions, Fig. 3 results as shown.

Since the equations are solved under no-load conditions, the voltage build up for both series and shunt connections are similar. The difference will arise only due the effect of rotor cage during the voltage build-up process as shown in Fig. 4, but in all cases the frequency of the generated voltage is 60Hz and peak no load voltage is 120V. It is seen that generators with rotor cage develop voltages gradually while a spontaneous voltage build-up with voltage overshoots occurs in the cageless machines.

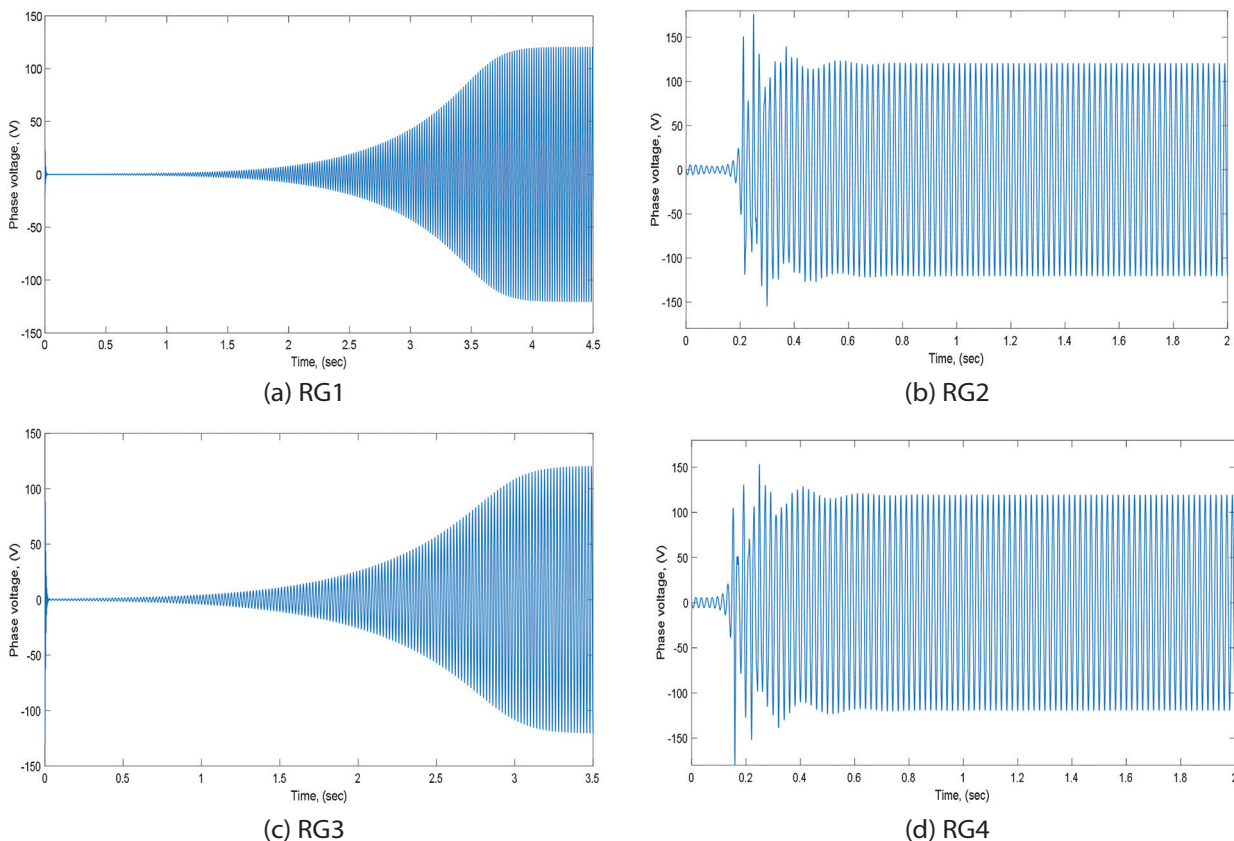


Figure 3. Voltage build up on phase A under no-load condition

5. TRANSIENT PERFORMANCE

The transient performance due to adding and removing of loads are now considered. The unloaded generator is operating at open circuit voltage when load is added at $t=7.0$ secs. This leads to some voltage oscillations and then the voltage settles to a new voltage value less than the open circuit value. Then the load is removed at $t=9.0$ secs. For a pure resistance (upf) load, the inductive loads are set to zero in the respective model equations. The variation in the values of voltages is used for assessing the loss of load performances. Results of plots from the four generator topologies are presented in Fig. 4 for unity power factor

tor (pure resistive) loads. Fig. 5 shows the corresponding results obtained for an RL load of 0.85 power factor.

In the four topologies considered, addition of load results less and quicker damping than when the same load is disconnected. This is akin to what happens in field-excited alternators [28]. Expectedly, machines with cage possess better damping characteristics than the cageless types. For the same magnitude of load, the resistive load yields a lower voltage drop but with more oscillations than the RL loads (Figs. 4a and 5a). Series connected machines are faster in damping than shunt-connected machines for similar loading conditions.

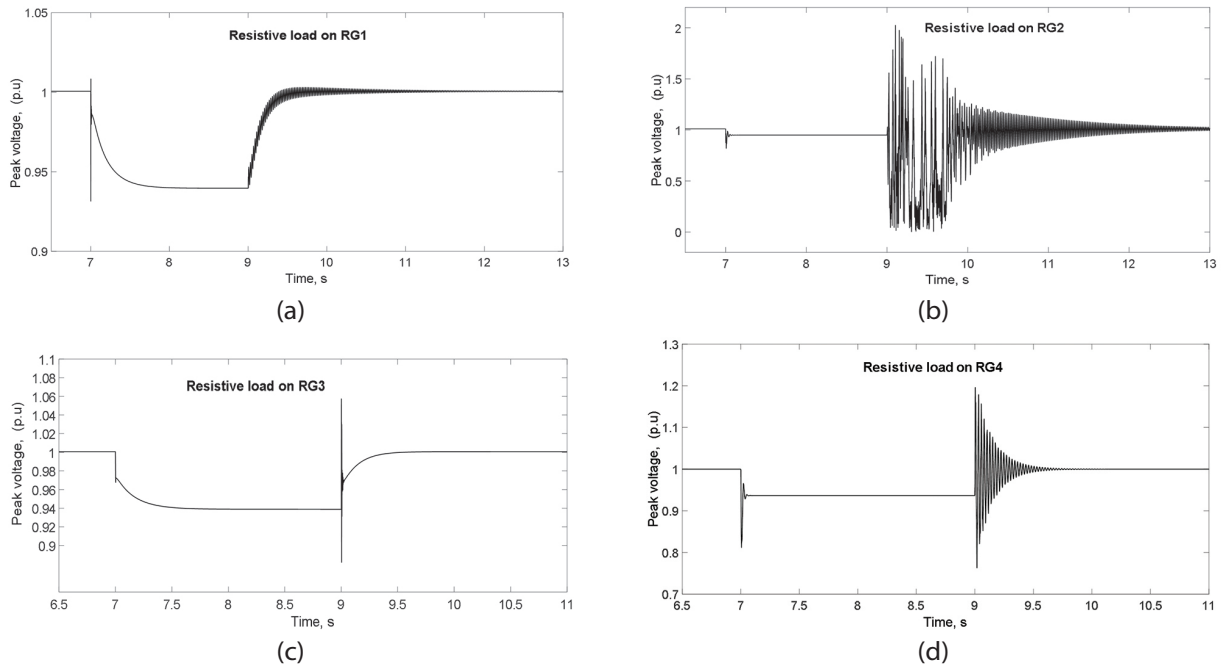


Fig. 4. Transient responses of phase A voltage following sudden switching and loss of pure resistive load on the generators

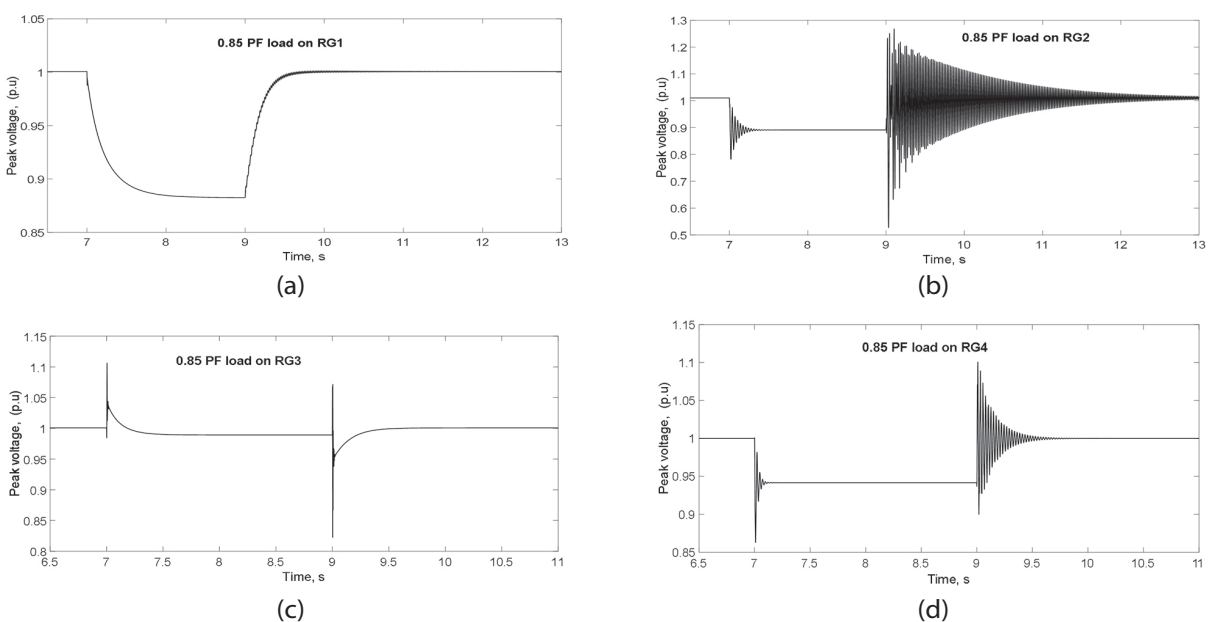


Fig. 5. Transient responses of phase A voltage following sudden switching and loss of 0.85 power factor load on the generators

Series generators (RG3 and RG4) are more capable of withstanding load transients than shunt generators (RG1 and RG2) for similar load combinations. In fact, the series generator with cage offers better damping of transients than the corresponding shunt machine.

6. VOLTAGE REGULATION CHARACTERISTICS

The non-linearity of L_d , L_q and R_c necessitates that the steady-state solution of self-excited generators be performed numerically. This can be accomplished by setting all the time derivatives of state variable to zero and then using either look up tables or quasi-Newton's iterative techniques (or similar algorithms) to solve the resulting set of equations with preset initial conditions. These have been reported severally in literature [2–5, 10, 11, 13] and it is not the intention of this paper to repeat such a study. In this study, the regulation is studied by loading each of the four generator types gradually from no load to full load and the peak value of the generator voltage observed. A full load point here is determined as the point where the voltage collapses or self-excitation is lost. Table 2 is obtained for each generator topology by gradual loading from no load to full load determined by a load angle of 43° and shows the results for three different load power factors.

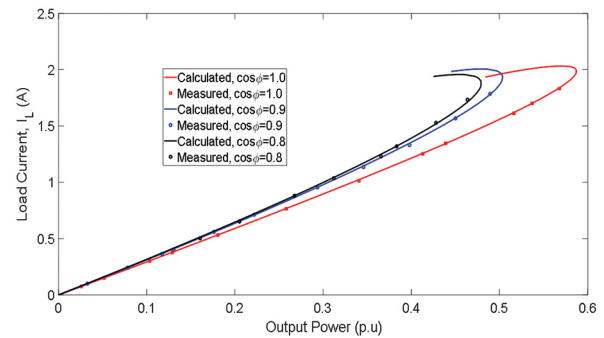
The regulation shows that the shunt generators are better power converters than series generators since they can deliver more power at every power factor. This is because for a series machine, the sum of leakage plus load inductance should not be higher than the value of the saturated q-axis inductance, otherwise system will fail to excite or if already excited, voltage collapse will result. All the topologies have their best voltage regulation and power carrying capacity at unity power factor.

Table 2. Maximum power deliverable and voltage regulation at $\delta=43^\circ$ electrical for the generators at different load power factors.

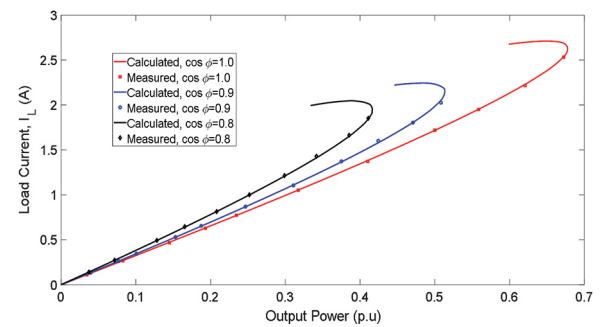
Load Power factor \rightarrow Gen. Topology \downarrow		$\cos\phi=0.8$	$\cos\phi=0.9$	$\cos\phi=1.0$
RG1&RG2 (shunt connection)	Max. Power (p.u)	0.466	0.579	0.691
	Voltage Reg. (%)	2.44	2.39	1.97
RG3&RG4 (series connection)	Max. Power (p.u)	0.453	0.511	0.589
	Voltage Reg. (%)	2.50	2.45	2.05

Measured and calculated values of load current plotted against the output power are shown in Fig. 6 for both the series and shunt generator. These again re-emphasizes that shunt generators are better in terms of energy-conversion than series generators. Fig. 7 shows the voltage of the SERG under steady state conditions with loads of differing power factors. The shunt generator exhibits better voltage regulation and is seen to be capable of supporting more load than the series

machine at all corresponding power factors. However, series generators are less sensitive to power factor variations than shunt generators

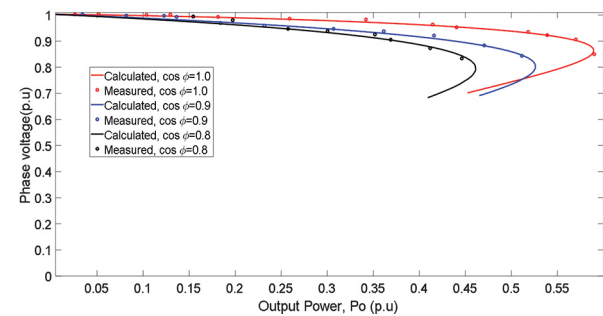


(a) Series generators



(b) Shunt generators

Fig. 6. Load current plots



(a) Series generators

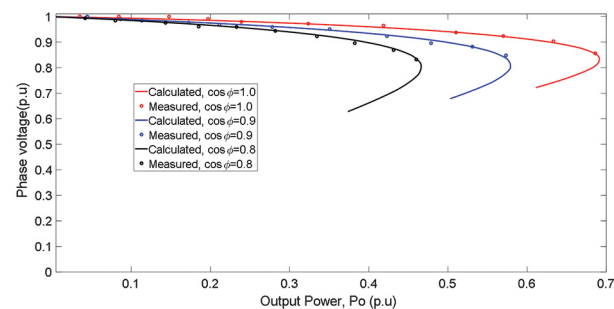


Fig. 7. Power vs Voltage plots

7. CONCLUSIONS

The seeming advantage of series connection is that both the excitation and load current flow in the connected load while only a fraction of the current in the

stator winding is available to the load for shunt connection. It would have then been easily presumed that series connection can support a larger load than shunt connection since all the winding current flows through the load, but results presented here has shown otherwise. This is because, for series connection, the load must be matched to the generator such that it fulfils the condition for self-excitation before a voltage can be produced.

The dynamics and voltage regulation characteristics of four stand-alone reluctance generator configurations has been studied. Expectedly, machines with damping cage are better in delivering a better voltage waveform than the cageless machine, even though the later are faster in excitation. Shunt generators are also seen to be better in terms of both voltage regulation and power conversion ability. Following sudden addition and loss of loads, series generators offer the advantage of having better damping characteristics than the shunt generators. Loss of load for cageless generator with shunt connection (RG2) leads to very large drop in voltages which may not be tolerable in some applications. Based on these results, the study recommends that a cageless rotor machine be used with series connection while a shunt generator requires a rotor cage.

8. REFERENCES

- [1] F. E. Abdel-Kader, "The reluctance machine as a self-excited reluctance generator", *Electrical Machines and Power Systems*, Vol. 10, No. 2 & 3, 1985, pp. 141-148
- [2] Y. H. A. Rahim, A. L. Mohamadien, A. S. Al Khalaf, "Comparison between the steady-state performance of self-excited reluctance and induction generators", *IEEE Transactions on Energy Conversion*, Vol. 5, No. 3, 1990, pp. 519-525.
- [3] M. N. Nagrial, M. A. Rahman, "Operation and characteristics of self-excited reluctance generator", *Record of IEEE-IAS 1988, Annual meeting, Part I*, pp.55-58.
- [4] T. F. Chan, "Steady-state analysis of a three-phase self-excited reluctance generator", *IEEE Transactions on Energy Conversion*, Vol. 7 No. 1 1992, pp. 223-230.
- [5] Y. Wang, N. Bianchi, "Analysis of Self-Excited PM-Assisted Reluctance Generators," *IEEE Transactions on Energy Conversion*, Vol. 33, No. 2, 2018, pp. 877-885.
- [6] A. L. Mohamedein, Y. H. A. Rahman, S. S. H. Khalaf, "Steady state performance of self-excited reluctance generators", *Proceedings of IEE*, Vol. 136, pt. B, 1990, pp. 293-298.
- [7] O. Ojo, "Limit-Cycle and small-signal dynamics of self-excited synchronous reluctance generator", *Record of IEEE-IAS 1988, Annual meeting*, pp.244-248.
- [8] L. Wang, Y.-S. Wang, "Dynamic performance and minimum loading effects of an isolated self-excited reluctance generator", *IEEE Power Engineering Society Winter Meeting*, Vol. 1, No. 31, 1999, pp. 13-18.
- [9] I. Boldea, X. Fu, S. A. Nasar, "High performance reluctance generator", *Proceedings of IEE, Part B*, Vol. 140, 1993, pp. 126-130.
- [10] N. Ben-Hail, R. Rabinovici, "Three-phase autonomous reluctance generator", *IEE Proceedings – Electric Power Applications*, Vol. 148, No.5, 2001, pp. 438-442.
- [11] A. Vagati, M. Pastorelli, F. Scapino, G. Franceschini, "Impact of Cross Saturation in Synchronous Reluctance Motors of the Transverse-Laminated Type", *IEEE Transactions on Industry Applications*, Vol. 36, No. 4, 2000, pp. 1039-1046.
- [12] E. S. Obe, L. U. Anih, "Influence of rotor cage on the performance of synchronous reluctance generators", *Electric Power Components and Systems*, Vol. 38, 2010, pp. 960-973.
- [13] M. Ibrahim, P. Pillay, "The Loss of Self-Excitation Capability in Stand-Alone Synchronous Reluctance Generators", *IEEE Transactions on Industry Applications*, Vol. 54, No. 6, 2018, pp. 6290-6298.
- [14] S. Guha, N. C. Kar, "Saturation Modeling and Stability Analysis of Synchronous Reluctance Generator", *IEEE Transactions on Energy Conversion*, Vol. 23, No. 3, 2008, pp. 814-823.
- [15] E. Hoffer, R. H. Moncada, B. J. Pavez-Lazo, J. A. Tapia, L. Laurila, "Calculation of a Current Vector Trajectory for Enhanced Operation of Synchronous Reluctance Generators Including Saturation", *IEEE Transactions on Industrial Electronics*, Vol. 70, No. 2, 2023, pp. 1197-1204.
- [16] Y. Wang, N. Bianchi, "Investigation of Self-Excited Synchronous Reluctance Generators", *IEEE Transactions on Industry Applications*, Vol. 54, No. 2, 2018, pp. 1360-1369.

- [17] S. S. Maroufian, P. Pillay, "Self-Excitation Criteria of the Synchronous Reluctance Generator in Stand-Alone Mode of Operation", *IEEE Transactions Industry Applications*, Vol. 54, No. 2, 2018, pp. 1245-1253.
- [18] Y. Wang, N. Bianchi, "Modeling and Investigation of Self-Excited Reluctance Generators for Wind Applications", *IEEE Trans. Industry Applications*, Vol. 55, No. 6, 2019, pp. 5809-5817.
- [19] J. S. Sedky, H. M. Yassin, H. H. Hanafy, F. Ismail, "Voltage and frequency control of standalone wind-driven self-excited reluctance generator using switching capacitors", *Journal of Electrical Systems and Information Technology*, Vol. 8, No. 66, 2021, pp. 1-24.
- [20] H. Guo, X. He, J. Xu, W. Tian, X. Ding, L. Ju, D. Li, "Design and analysis of a novel hybrid cooling method of high-speed high-power permanent magnet assisted synchronous reluctance starter/generator in aviation applications", *Chinese Journal of Aeronautics*, Vol. 36, No. 3, 2023, pp 285-302.
- [21] R. R. Kumar, S.K. Singh, R. K. Srivastava, R. K. Saket, "Dynamic reluctance air gap modeling and experimental evaluation of electromagnetic characteristics of five-phase permanent magnet synchronous generator for wind power application", *Ain Shams Engineering Journal*, Vol. 11, No. 2, 2020, pp. 377-387.
- [22] O. D. Momoh, T. O. Ajewole, M. N. O. Sadiku, "Investigating the effects of faulty induction machine in distribution systems", *Scientific African*, Vol. 9, 2020.
- [23] I. O. Ozioko, N. S. Ugwuanyi, A. O. Ekwue, C. I. Odeh, "Wind energy penetration impact on active power flow in developing grids", *Scientific African*, Vol. 18, 2022.
- [24] H. A. Rahim, J. E. Fletcher, N. E. A. M. Hassanain, "Performance analysis of salient-pole self-excited reluctance generators using a simplified model", *IET Renewable Power Generation*, Vol. 4, No. 3, 2010, pp. 253-260.
- [25] A. I. Alolah, "Capacitance requirements for three phase self-excited reluctance generators", *IEE Proceedings - Generation, Transmission and Distribution*, Vol. 138, No. 3, 1991, pp. 193-198.
- [26] T. R. Ayodele, A. S. O. Ogunjuyigbe, B. B. Adetokun, "Optimal capacitance selection for a wind-driven self-excited reluctance generator under varying wind speed and load conditions", *Applied Energy*, Vol. 190, 2017, pp. 339-353.
- [27] A. S. O. Ogunjuyigbe, T. R. Ayodele, B. B. Adetokun, "Steady state analysis of wind-driven self-excited reluctance generator for isolated applications", *Renewable Energy*, Vol. 114, Part B, 2017, pp. 984-1004.
- [28] Y. Zhan, K. Kong, G. Xu, J. Kang, H. Zhao, "Analysis of Damper Transient Currents in Salient-Pole Synchronous Generator with Skewed Armature Slots Considering Interbar Currents", *IEEE Transactions on Industry Applications*, Vol. 55, No. 1, 2019, pp. 336-343.

Design and analysis of a new multi-level inverter topology with a reduced number of switches and controlled by PDPWM technique

Case Study

Fatima Chakir

IESI Laboratory, Department of Electrical Engineering, Faculty of Science and Technology, ENSET Mohammedia, Hassan II University Of Casablanca, Morocco, fatima.chakir5-etu@etu.univh2c.ma

Abdelmounime EL Magri

IESI Laboratory, Department of Electrical Engineering, Faculty of Science and Technology, ENSET Mohammedia, Hassan II University Of Casablanca, Morocco, magri_mounaim@yahoo.fr.

Rachid Lajouad

IESI Laboratory, Department of Electrical Engineering, Faculty of Science and Technology, ENSET Mohammedia, Hassan II University Of Casablanca, Morocco, rlajouad@gmail.com.

Mohamed Kissaoui

IESI Laboratory, Department of Electrical Engineering, Faculty of Science and Technology, ENSET Mohammedia, Hassan II University Of Casablanca, Morocco, kissaouimed@gmail.com

Mostafa Chakir

CED-ST, LESSI, Faculty of Sciences Dhar el Mehraz, Sidi Mohamed Ben Abdellah University, BP 1796, Fez-Atlas, 30003 Fez, Morocco, mostafa.chakir@usmba.ac.ma

Omar Bouattane

IESI Laboratory, Department of Electrical Engineering, Faculty of Science and Technology, ENSET Mohammedia, Hassan II University Of Casablanca, Morocco, o.bouattane@enset-media.ac.ma.

Abstract – With their many advantages, including low power dissipation in power switches, low harmonic content, and reduced electromagnetic interference (EMI) from the inverter, multilevel converter (MLI) topologies are becoming more and more in demand in high and medium power applications. This paper introduces a novel multi-level symmetric inverter topology with adopted control. The objectives of this article are to architecturally define the positions of the various switches, to choose the right switches and to propose an inverter control strategy that will eliminate harmonics while producing the ideal output voltage/current. By using fewer switching elements, fewer voltage sources, and switches with a total harmonic content (THD) which reduces losses and a drop in minimum voltage (V_{strssj}), the proposed topology is more efficient than conventional inverters with the same number of levels. The new topology will be demonstrated using a seven-level single-phase inverter. For various modulation indices, MATLAB-SIMULINK is used to study and validate the topology.

Keywords: Approximate controls, Inverter multi-level, Pulse with modulation (PWM), Total harmonic distortion (THD)

1. INTRODUCTION

Renewable energy is an inexhaustible and freely available source of power. In addition to being viewed as ecologically clean and limitless on a human scale, sources like the sun's rays (solar energy), the wind (wind energy), and the kinetic energy of water (hydraulic energy) are used to produce so-called renewable energy.

Renewable energy is a cost-free and limitless energy source. Such resources as the solar, wind and kinetic energy of water are considered to be ecologically sound and, in theory, limitless on a human scale [1].

A good angle of inclination of the solar collector improves productivity and yield [2]. Since the energy produced by renewable sources is continuous and un-

stable [3], this causes instability in the production of electricity and contributes to the fidelity and quality of the electrical system. DC/AC converters are crucial to solve these problems [4].

The two most well-known types of classical structures are two-level inverters and three-level or modified square wave inverters. The latter, however, can be applied to specific applications [5]. Due to their numerous drawbacks, such as their high harmonic rate, degraded fundamental voltage, large size, and extremely high rated switching voltage, they are not recommended for new projects [6, 7]. These problems were addressed by the design of multi-level inverters (MLI).

Multi-level inverters have received a lot of attention in recent years as a potential solution for high-power continues /alternative conversion applications. A PWM is a structure that connects various DC current or voltage sources (produced by a DC source, capacitors), power switches (MOSFET, IGBT), and results in the creation of a stepped waveform. At the terminals of the power switches, the voltage drops are smaller than the converter voltage as a whole[8]. Comparatively to conventional two-level inverters, multi-level inverters have a better harmonic profile of their output waveform [9, 10]. Moreover, fault-tolerant operation is possible [11]. The complexity and cost of the PWM configuration remain the main drawbacks due to the abundance of switching devices [12]. PWMs are frequently used for high and low power applications. The NPC(neutral-point clamped), the CHB (cascaded H-bridge), and the FC (flying capacitor inverter) are, however, the most well-known multilevel inverter topologies [13-17].

The new topology incorporates the same number of switches as the CHB topology which requires fewer switching elements compared to capacitor (CF) or diode (NPC) topologies, overriding the harmonic ratio of the voltage generated by the new topology (17.15%) is lower than that generated by the CHB (THBCHB = 18.3%) by adopting the same control technique (PDPWM) [18]. The new topology with its ability to generate an output voltage with a reduced harmonic content (compared to CHB) and a minimum number of switches (compared to FC and NPC) for the same number of levels, presents a good alternation by the reduction of switching losses since the number of the latter is reduced, its manufacturing cost and its size thus compared to the conventional converter, which attracts many researchers and industrial collaborators [7].

The article is structured as follows: Section II presented a generalized configuration of the proposed topology with mathematical formulations; this section also describes the operation mode of a single-phase, 7-level inverter. The modulation method (PWM) is introduced in section III. The results of the simulation are then presented in section IV. Section V compares the proposed topology and the traditional topologies in terms of the number of interrupters and THD. Conclusions are drawn in the final section.

2. THE SUGGESTIVE TOPOLOGY.

In this section, the structure of the new topology and its working principle for a single-phase seven-level PWM are presented. Formulas for voltage stresses across switches, output voltage, etc. are also presented.

2.1. THE NEW TOPOLOGY'S GENERAL STRUCTURE

The generalized structure of the new single-phase inverter topology is provided by (Fig. 1); it integrates n DC voltage sources connected in series by power switches.

The switches can be MOSFET or IGBT transistors with an antiparallel diode, connect the input sources E_j (where $j = 1$ to n) between them in order to produce the currents $i_{1j}(t)$ and $i_{2j}(t)$. IGBTs and antiparallel diodes are used in this inverter structure to model the power switches. In addition to two latch interrupts (S_1, S_2), complementary switches are identified by Q_{1j}, Q_{2j} (where $j = 1$ to n). $v_{1j}(t)/v_{2j}(t)$ (where $j=1$ to n) represents the voltage across the terminals of each switch. The symbols for load current and voltage are $i_{ab}(t)$ and $V_{ab}(t)$, respectively.

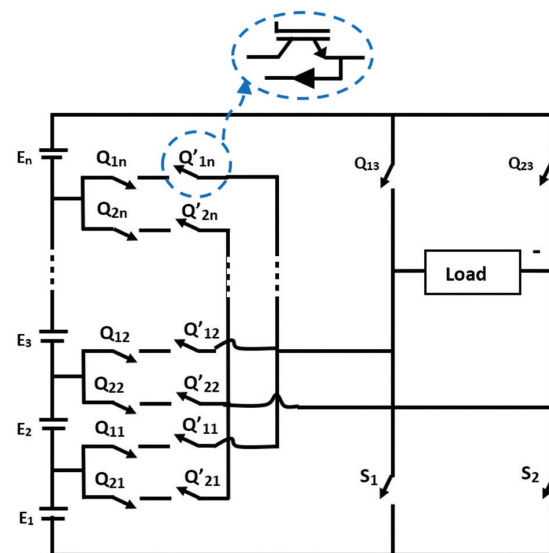


Fig. 1. The general proposed topology of the single-phase inverter

The new proposed symmetric topology of multi-level (SMLI) and mode of operation are shown in Fig. 2 and Table 1, respectively. For a 7-level output, it is necessary to use three voltage sources (E_1, E_2 and E_3) and twelve pairs of unidirectional switches, which are made up of IGBT transistors combined with an anti-parallel diode. The proposed topology results in a maximum voltage of $3E$ for the 7-level PWM:

$$(E_1 + E_2 + E_3 = 3E).$$

Fig. 2 clearly shows that the combination of switches (Q_{13}, Q_{12}), (Q_{13}, Q_{11}), (Q_{13}, S_1), (Q_{12}, Q_{11}), (Q_{11}, S_1), (Q_{12}, S_1), (Q_{23}, Q_{22}), (Q_{23}, Q_{21}), (Q_{23}, S_2), (Q_{22}, Q_{21}), (Q_{21}, S_2) and (Q_{22}, S_2) in order to avoid short-circuiting of sources; those interrupter shouldn't be activated in the same time.

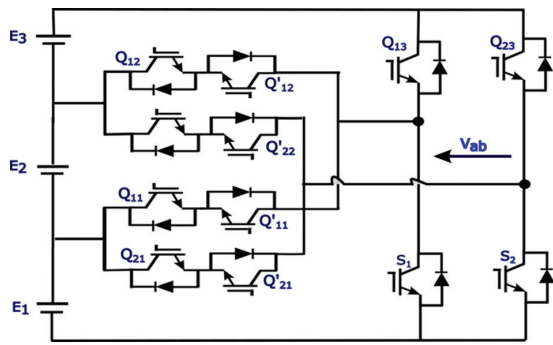


Fig. 2. The proposed topology for a 7-level symmetrical inverter

Table 1. The output voltage values Vab for different status of interrupters (F:OFF/ N:ON)

Level	State of switches								Vab
	Q ₁₃	Q ₁₂	Q ₁₁	S ₁	S ₂	Q ₂₃	Q ₂₂	Q ₂₁	
1	F	F	F	O	O	F	F	F	0
2	F	F	O	F	O	F	F	F	E
3	O	F	F	F	F	F	F	O	2E
4	O	F	F	F	O	F	F	F	3E
5	O	O	O	F	F	O	F	F	-E
6	F	F	F	O	F	F	O	F	-2E
7	F	F	F	O	F	O	F	F	-3E

A switch can present in more than one level structure to the output voltage in the single-phase inverter structure that is being proposed.

To achieve higher output voltage levels, the new inverter structure can be expanded.

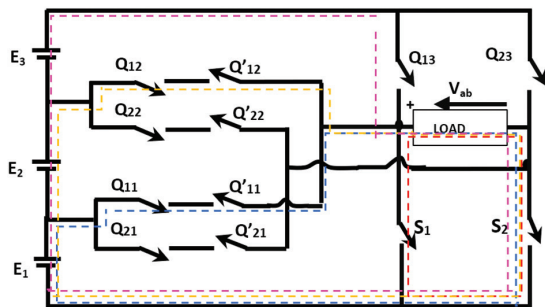


Fig. 3. Output voltage under operating mode of proposed topology: Vab=0; Vab=E; Vab=2E; Vab=3E;

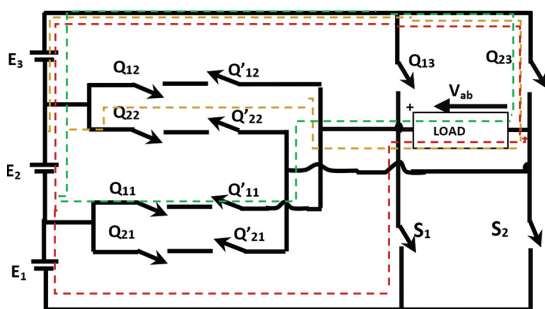


Fig. 4. Output voltage under operating mode of proposed topology: Vab= -E; Vab= -2E; Vab= -3E;

2.3. MATHEMATICAL FORMULATIONS

Based on the Fig.1, the switches Q_{1j} and Q_{2j} ($j=1$ to n) function in a complementary mode, the switching functions of Q_{1j}/Q_{2j} which are described as follows:

$$Q_{1j} = \begin{cases} 1, & \text{if "ON"} \\ 0, & \text{if "OFF"} \end{cases} \quad (1)$$

$$Q_{2j} = \begin{cases} 1, & \text{if "ON"} \\ 0, & \text{if "OFF"} \end{cases} \quad (2)$$

Then the load voltage $v_{ab}(t)$ can be formulated using the nodal voltages $v_a(t)$ and $v_b(t)$ as:

$$V_{ab}(t) = V_a(t) + V_b(t) \quad (3)$$

Or:

$$V_a(t) = Q_{1j} * j * E \quad (4)$$

And:

$$V_b(t) = Q_{2j} * j * E \quad (5)$$

By using (3), (4) and (5):

$$V_{ab}(t) = Q_{1j} * j * E - Q_{2j} * j * E \quad (6)$$

Furthermore, the number of levels generated by the proposed typology is given by

$$N = 2 * n + 1 \quad (7)$$

Or:

n : is the number of continuous sources,

N : is the number of levels of the stepped output

The maximum voltage is given by:

$$V_{(max)} = n * E \quad (8)$$

3. THE TECHNIQUES OF PWM MODULATION

The goal of modulation techniques is to create switch drive pulses that generate a voltage close to a sinusoidal shape at the output of the multilevel inverter (MLI) [18-19]. PWM techniques are used to reduce the harmonic level of the output wave and they can be classified as follows:

- APOD: Alternative Phase Opposition Disposition.
- PD: Phase Disposition Fig.5.
- POD: Phase Opposition Dispositions.

According to the literature, the PDPWM technique generates a better harmonic profile than the other methods (PDPWM & APODPWM) with a CHB inverter [18-21], therefore, to validate the usefulness of the proposed topology, we will adopt the same technical (PDPWM).

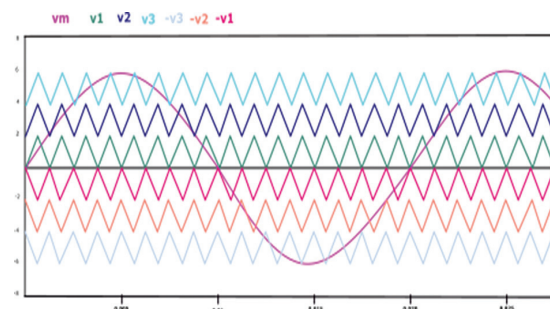


Fig. 5. The Phase Disposition techniques of PWM modulation

4. SIMULATION AND RESULTS

The MATLAB/Simulink tool was used; to examine the operation and inspected the results of the new topology. The latter is tested with the configurations already defined in Table 1 with the appropriate PDPWM command. The test parameters of the new 7-level DC/AC converter are given by Table 2. In this paper; two types of control are used: the full wave control to test the operation of the proposed topology -in the first time- and the PDPWM control in the second time. The aim is improving the performance of the output of our inverter (minimized the harmonic rate of the voltage and current produced).

Table 2. Simulation parameters

Input DC Voltage (V)	100
Charge Resistance R (Ω)	100
Charge Inductance L (H)	1e-1
Carrier signals (V)	100
carrier frequency (KHz)	4
Reference signal (V)	230
signal frequency (Hz)	50

a. Full Wave Control

The control of the switches are fixed according to the chronology given prior in the Table 1. The output voltage wave and its total harmonic distortion realized using FFT analysis of MATLAB/Simulink are presenting in (Fig. 5).

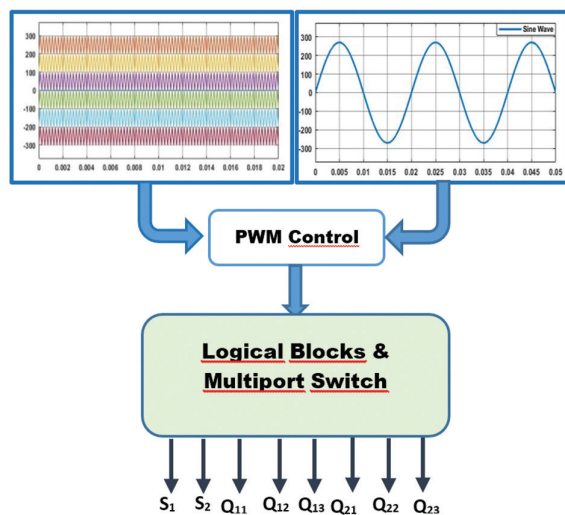
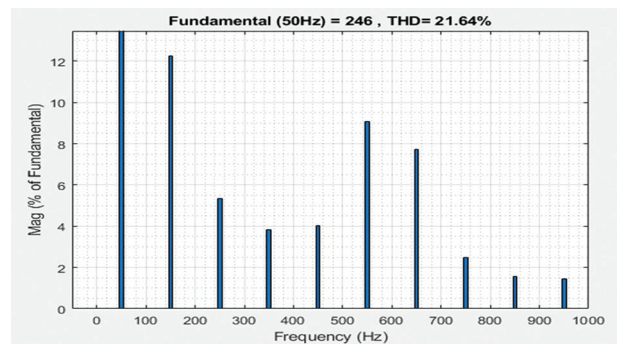
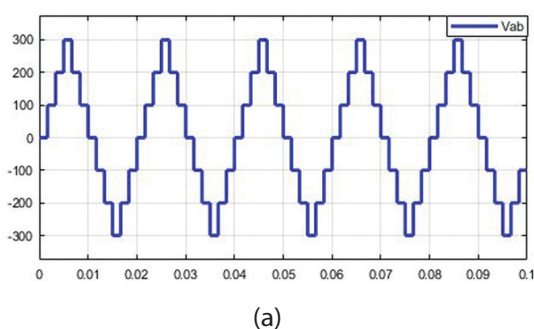


Fig. 6. Synoptic of the command adopted



(b)

Fig. 7. (a) Proposed seven- level output voltage waveform; (b) FFT analysis of voltage single phase MLI.

The THD of voltage output is 21.64 % with 3rd, 5th, 7th and 9th harmonics 12%, 5.6%, 3.8% and 4.03% respectively.

b. The multi-carrier PWM control.

References (sine waves) are compared to carriers (triangular waves). With F_{sin} and F_{tag} being the sine waveform and triangle waveform frequencies, respectively.

At first the values of V_{sin} , V_{tag} are: 230v, 100 v respectively; these last ones will be changed to obtain the different modulation index values. The carrier frequency used in the simulation is 10KHz.

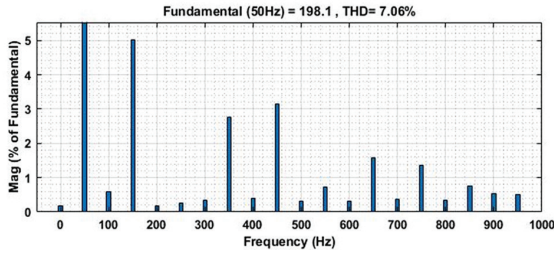
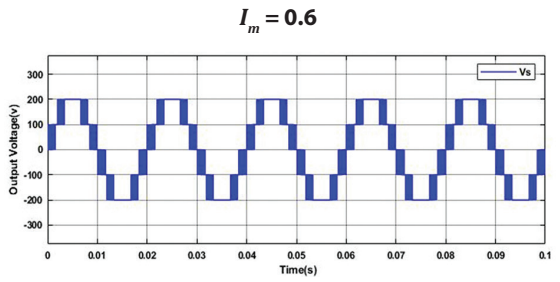
V_{tag} is the amplitude of carrier signal;
 V_{sin} is the amplitude of reference signal;

Then I_m is the modulation index is defined as:

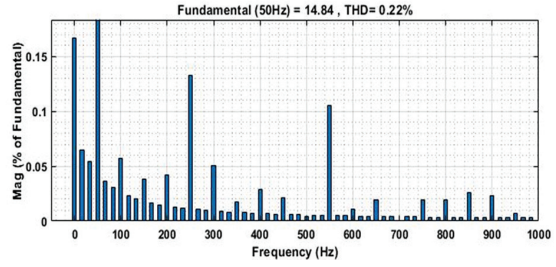
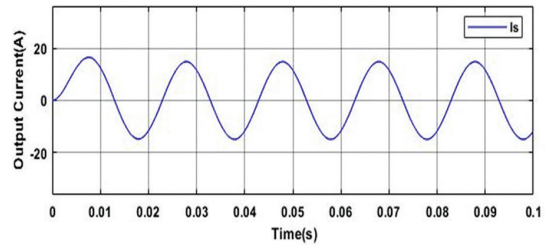
$$I_m = \frac{V_{sin}}{3 * V_{tag}} \quad (9)$$

The design goals of PWM inverters are to reduce the THD in the output current and voltage. To test the rate of harmonics, the simulations are carried out with an RL with ($L=1e^{-1}$) load and without using a filter. (Fig. 8) shows the current and voltage waveforms of our 7-level inverter as well as their THD for modulation indices from 0.6 to 1 with the same amplitude. Based on an analysis of the voltage and current waveforms exhibited by the new topology, the most appropriate output voltage profile (voltage and current close to the sinusoid) are given by a modulation index $I_m=0.8$ (THDV=1.24% and THDI = 0.22% respectively). therefore, for the same number of output levels, we can see that the new topology generates a lower rate of harmonics than the topology proposed by [18], we constat that the new proposed topology generate a rate of current harmonics lower than the topology proposed by N.Motaparthi and all. in the two cases of study (without filter THDI=6.47% , with THDI= 2.89% filter) [22].

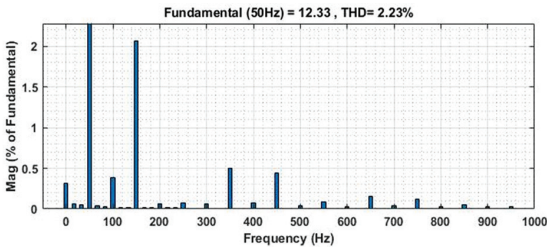
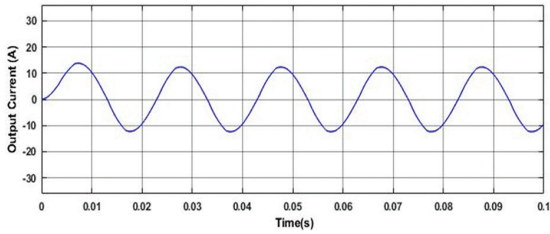
The Table 3 below; illustrates how the THD decreases with decreasing modulation index, resulting in fewer output voltage levels ($I_m=0.6$ generates 5 levels) Fig.8.a.



(a)

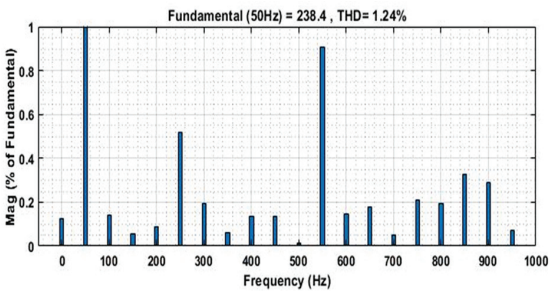
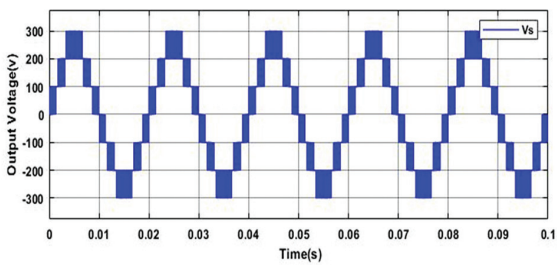


(b')



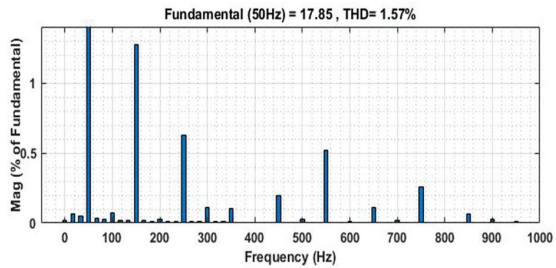
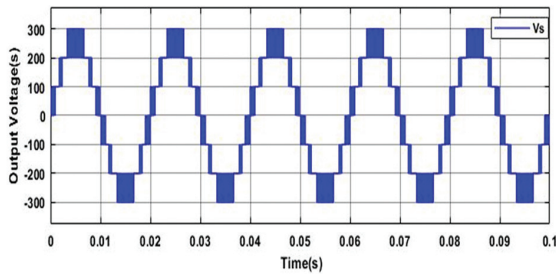
(a')

$I_m = 0.8$

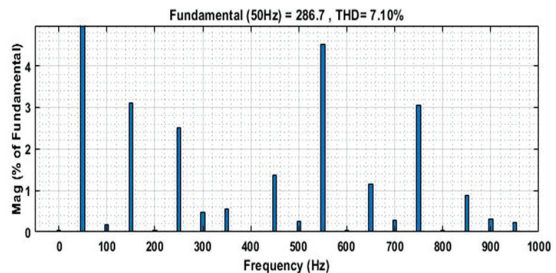
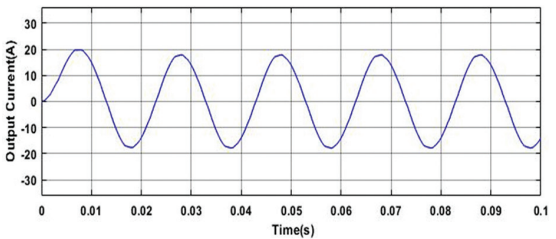


(b)

$I_m = 1$



(c)



(c')

Fig. 8. Voltage, current waveforms generated by the new 7-level PWM and their THD at different modulation index: (a); (a') $I_m = 0.6$, (b); (b') $I_m = 0.8$, (c); (c') $I_m = 1$

The design goals of PWM inverters are to reduce the THD in the output current and voltage. To test the rate of harmonics, the simulations are carried out with an RL_ with ($L=1e^{-1}$) load and without using a filter. (Fig. 8) shows the current and voltage waveforms of our 7-level inverter as well as their THD for modulation indices from 0.6 to 1 with the same amplitude. Based on an analysis of the voltage and current waveforms exhibited by the new topology, the most appropriate output voltage profile (voltage and current close to the sinusoid) are given by a modulation index $I_m=0.8$ (THDV=1.24% and THDI = 0.22% respectively). therefore, for the same number of output levels, we can see that the new topology generates a lower rate of harmonics than the topology proposed by [18], we constant that the new proposed topology generate a rate of current harmonics lower than the topology proposed by N.Motaparthi and all. in the two cases of study (without filter THDI=6.47% , with THDI= 2.89% filter) [22].

The Table 3 below; illustrates how the THD decreases with decreasing modulation index, resulting in fewer output voltage levels ($I_m=0.6$ generates 5 levels) Fig. 8. a.

Table 3. The variation of THD of voltage waveform as a function of different modulation index

Modulation Index	THDI% (THD of Current)	THDV% (THD of voltage)
0.6	2.23	7.06
0.7	0.44	2.23
0.8	0.22	1.24
0.9	0.69	3.81
1	1.57	7.1

5. COMPARATIVE STUDY

A comparison is presented in Table 4 between the new proposed topology , the Two-levels topology and three classics topologies: CHB, FC, and NPC [23-24].

Moreover, equation (7) gives the number of levels generated by the proposed topology.

The comparison between the new inverter topology, the two-level structure and the classic topologies: FC, CHB and NPC is given by Table 4, it is clear that the proposed symmetrical topology requires the same number of switches as the topology CHB presented in the article [23-24]; to generate particular output levels, on the other hand, it requires the minimum of components compared to FC and NPC topologies, as shown in table 4. we see that the new symmetrical structure does not need a flying capacitor and a clamping diode, which has an advantage of the latter over FC, NPC and the topologies presented in other literatures [6] because it reduces the complexity of the control circuit and avoids the balancing problem caused by the presence of the capacitors presented in the NPC and FC topologies. Looking at the THDs we notice that the topology has the lowest harmonic rate.

Therefore, the new proposed topology has proven to be the most efficient in producing voltage and current with the lowest THD while integrating the minimum number of switches compared to conventional topologies.

Table 4. Comparative study between the new proposed topology, two levels inverter and the classics structures according to the number of components

Components	Topologies				
	New Topology	NPC-MLI	FC-MLI	CHB-MLI	Two levels
DC Sources	(N-1)/2	1	1	(N-1)/2	1
Switches	2*(N-1)	2*(N-1)	2*(N-1)	2*(N-1)	2
Bus capacitors	0	(N-1)	(N-1)	0	1
Clamping diodes	0	2*(N+1)	0	0	0
Flying capacitors	0	0	(N+1)	0	0
Modulation technique	PD PWM	PD PWM	PD PWM	PD PWM	SVP PWM
THD(%) of current	0.22%	1.26%	4.22%	0.26%	6.02%
Total components for 7 levels inverter (N=7)	15	35	27	15	-

6. CONCLUSION

The purpose of this article is to present a new multilevel inverter topology that can be used in engineering applications. In this article, a single-phase 7-level inverter was used to validate and demonstrate the efficiency of the model. Researchers found that increasing output levels reduced total harmonic distortion (THD), but also increased losses and increased cost of the inverter as more switches were required. Therefore, our study focuses on developing a structure with a trade-off between the number of switches used and the number of output levels. Modulation index (I_m) variation influences inverter performance; for this reason the results of the simulation are presented for a different modulations indexes; varying from 0.6 to 1 for a good explanation of the results. Based on the FFT analysis of the output quantities (voltage/current) and comparing with previous studies, it appears that the topology produces a good voltage profile (THDV = 1.24%) for $I_m = 0.8$ without using additional filter. For the same number of output levels, the proposed topology appears to have the lowest THD for output voltage and current, therefore requires fewer components and therefore has less switching losses and does not include capacitors which exhibit the disadvantage of voltage balancing for classical topology. Consequently, the proposed topology is a good alternative for engineer applications.

Acknowledgements

This article is included in the UPISREE research project funded and supported by the organization Research in Solar Energy and New Energies (IRESEN).

7. REFERENCES:

- [1] S. M. Shaahid, "Review of research on autonomous wind farms and solar parks and their feasibility for commercial loads in hot regions", *Renewable and Sustainable Energy Reviews*, Vol. 15, No. 8, 2011, pp. 3877-3887.
- [2] X. Berisha, A. Zeqiri, D. Meha, "Determining the Optimum Tilt Angles to Maximize the Incident Solar Radiation - Case of Study Pristina", *International Journal of Renewable Energy Development*, Vol. 7, No. 2, 2018, pp. 123-130.
- [3] B. P. Chandran, A. I. Selvakumar, F. M. Mathew, "Integrating multilevel converters application on renewable energy sources—A survey", *Journal of Renewable and Sustainable Energy*, Vol. 10, No. 6, 2018, p. 065502.
- [4] K. Dhineshkumar, C. Subramani, A. Geetha, C. Vimala, "Performance analysis of PV powered multilevel inverter", *International Journal of Electrical and Computer Engineering*, Vol. 9, No. 2, 2019, p. 753.
- [5] W. Phetphimoon, Y. Kongjeen, K. Bhumkittipich, "Design of High-Frequency ZVS Full-Bridge Power Converter for Photovoltaic Applications", *International Review of Electrical Engineering*, Vol. 16, No. 6, 2021, p. 497.
- [6] K. K. Gupta, S. Jain, "A Novel Multilevel Inverter Based on Switched DC Sources", *IEEE Transactions on Industrial Electronics*, Vol. 61, No. 7, 2014, pp. 3269-3278.
- [7] A. Dekka, A. Ramezani, S. Ounie, M. Narimani, "A New 5-Level Voltage Source Inverter", *Proceedings of the IEEE Applied Power Electronics Conference and Exposition*, Anaheim, CA, USA, March 2019, pp. 2511-2515.
- [8] S. F. Mekhamer, R. H. Shehata, A. Y. Abdelaziz, M. A. Al-Gabalawy, "Enhancing radial distribution system performance by optimal placement of DSTATCOM", *International Journal of Electrical and Computer Engineering*, Vol. 10, No. 3, 2020, p. 2850.
- [9] S. Choudhury, M. Bajaj, T. Dash, S. Kamel, F. Jurado, "Multilevel Inverter: A Survey on Classical and Advanced Topologies, Control Schemes, Applications to Power System and Future Prospects", *Energies*, Vol. 14, No. 18, 2021, p. 5773.
- [10] S. Raj, R. K. Mandal, M. De, A. K. Singh, "Nine-level inverter with lesser number of power semiconductor switches using dSPACE", *International Journal of Power Electronics and Drive Systems*, Vol. 13, No. 1, Mar. 2022, p. 39.
- [11] G. Buticchi, E. Lorenzani, G. Franceschini, "A Five-Level Single-Phase Grid-Connected Converter for Renewable Distributed Systems", *IEEE Transactions on Industrial Electronics*, Vol. 60, No. 3, 2013, pp. 906-918.
- [12] A. M. Al-Mahrouk, N. F. Mailah, M. A. Mohd Radzi, M. K. Hassan, "Systematic Review of Multilevel and Matrix Usage in Power Electronics: Circuit Types, System Taxonomy, Applications and Recommendations", *International Review of Electrical Engineering*, Vol. 15, No. 2, 2020, p. 108.
- [13] S. Debnath, J. Qin, B. Bahrani, M. Saeedifard, P. Barbosa, "Operation, Control, and Applications of the Modular Multilevel Converter: A Review", *IEEE Transactions on Power Electronics*, Vol. 30, No. 1, 2015, pp. 37-53.
- [14] D. Roy, M. Singh, "A Single Z-Source Network-Based 3-Level NPC Inverter Using a Novel Region Selection Approach of SVM", *International Review of Electrical Engineering*, Vol. 15, No. 5, 2020, p. 394.
- [15] M. Malinowski, K. Gopakumar, J. Rodriguez, M. A. Pérez, "A Survey on Cascaded Multilevel Inverters", *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 7, 2010, pp. 2197-2206.
- [16] K. Compala Lakshmiah, T. A. Raghavendiran, "A New Modified H-Bridge Multilevel Inverter with Multi Carrier PWM Technique for Speed Control of Induction Motor", *International Review of Electrical Engineering*, Vol. 13, No. 5, 2018, p. 365.
- [17] F. Gülpınar, F. Sarı, Y. Uzun, "Analysis of a Novel Four Level Flying Capacitor H - Bridge Converter", *International Journal of Renewable Energy Development*, Vol. 7, No. 1, 2018, p. 71.
- [18] C. Manivelan, "A Survey on Multilevel Inverter Topologies and Control Schemes with Harmonic Elimination", *Proceedings of the International Conference on Electrotechnical Complexes and Systems*, Ufa, Russia, October 2020, pp. 1-7.
- [19] W.-K. Choi, F. Kang, "H-bridge based multilevel inverter using PWM switching function", *Proceedings of the 31st International Telecommunications Energy Conference*, Incheon, South Korea, October 2009, pp. 1-5.
- [20] S. W. Shneen, F. N. Abdullah, D. H. Shaker, "Simulation model of single phase PWM inverter by using MATLAB/Simulink", *International Journal of Power Electronics and Drive Systems*, Vol. 12, No. 1, 2021, p. 212.

- [21] R. Taleb, M. Helaimi, D. Benyoucef, Z. Boudjema, "A comparative analysis of multicarrier SPWM strategies for five-level flying capacitor inverter", Proceedings of the 8th International Conference on Modeling, Identification and Control, Algiers, Algeria, November 2016, pp. 608-611.
- [22] N. Motaparathi, K. Kumar Malligunta, "Seven-Level Symmetrical Series/Parallel Multilevel Inverter with PWM Technique Using Digital Logic", International Journal of Electrical and Computer Engineering Systems, Vol. 12, No. 3, 2021, pp. 123-130.
- [23] R. Mali, N. Adam, A. Satpaise, A. P. Vaidya, "Performance Comparison of Two Level Inverter with Classical Multilevel Inverter Topologies", Proceedings of the IEEE International Conference on Electrical, Computer and Communication Technologies, Coimbatore, India, February 2019.
- [24] S. P. Gautam, "Novel H-Bridge-Based Topology of Multilevel Inverter With Reduced Number of Devices", IEEE Journal of Emerging and Selected Topics in Power Electronics, Vol. 7, No. 4, 2019, pp. 2323-2332.

Biomolecular Event Extraction using Natural Language Processing

Review Paper

Manish Bali

Department of Computer Science and Engineering,
Presidency University, Bengaluru,
Karnataka, India
balimanish0@gmail.com

S. P. Anandaraj

Department of Computer Science and Engineering,
Presidency University, Bengaluru
Karnataka, India
anandsofttech@gmail.com

Abstract – Biomedical research and discoveries are communicated through scholarly publications and this literature is voluminous, rich in scientific text and growing exponentially by the day. Biomedical journals publish nearly three thousand research articles daily, making literature search a challenging proposition for researchers. Biomolecular events involve genes, proteins, metabolites, and enzymes that provide invaluable insights into biological processes and explain the physiological functional mechanisms. Text mining (TM) or extraction of such events automatically from big data is the only quick and viable solution to gather any useful information. Such events extracted from biological literature have a broad range of applications like database curation, ontology construction, semantic web search and interactive systems. However, automatic extraction has its challenges on account of ambiguity and the diverse nature of natural language and associated linguistic occurrences like speculations, negations etc., which commonly exist in biomedical texts and lead to erroneous elucidation. In the last decade, many strategies have been proposed in this field, using different paradigms like Biomedical natural language processing (BioNLP), machine learning and deep learning. Also, new parallel computing architectures like graphical processing units (GPU) have emerged as possible candidates to accelerate the event extraction pipeline. This paper reviews and provides a summarization of the key approaches in complex biomolecular big data event extraction tasks and recommends a balanced architecture in terms of accuracy, speed, computational cost, and memory usage towards developing a robust GPU-accelerated BioNLP system.

Keywords: Biomolecular event extraction, natural language processing, text mining, machine learning, BioNLP shared task

1. INTRODUCTION

Medical literature is a vast repository for knowledge sharing that happens in the biomedical domain. With major advances in computational biology and allied scientific research, there is an explosion in the number of publications in this area [1]. Every day approximately three thousand research articles are getting published in biomedical journals. Considering just one database, say MEDLINE, there are 23,000,000 references with 40,000-50,000 getting added every day. For any researcher, this poses an enormous challenge to locate, manage and choose suitable literature in their domain. Thus, the automated mechanism to extract structured content (explicit knowledge) from unstructured text (implicit knowledge) as shown in (Fig. 1) is the need of the hour [2]. The information deluge is posing new

challenges as bio-databases, vocabularies and bio-ontologies encode only a small fraction of information. Curators are struggling to process scientific literature. The discovery of facts and events is crucial for gaining insights into biosciences, hence the need for text event mining. Artificial Intelligence and especially Machine Learning techniques such as NLP and TM tools have gained significant importance to curate large biological databases. This has led to the development of many new applications and search engines addressing various domains for mining databases [3-5].

Previous research has focused on the basic extraction of entities and identifying their links in reference knowledge bases [6, 7]. Few existing techniques though provide acceptable performance [8] for many applications. Off-late interest has arisen in biomedical

entities, for example, drug-protein, drug-drug, and protein-protein interactions [10,11] which have emerged as the most important entities due to several similar databases and their usage in systems biology.

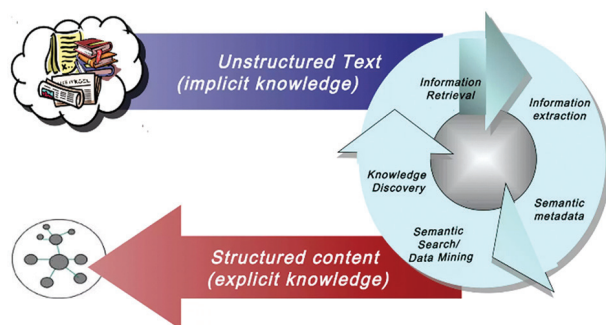


Fig. 1. Solving the biodata deluge using text mining (TM) [9]

NLP-based data-channeling process and their potential applications are many [12,13] as illustrated in (Fig. 2). Initially in the pipeline, anyone working in the biomedical NLP domain uses Information retrieval (IR) techniques which include tasks such as classification and retrieval of documents to select relevant articles. This process helps to reduce unwanted text or documents to only those that are of interest to the researcher. Next in the pipeline, Information extraction (IE) techniques are used to primarily identify any named entity and/or extract any event that is of researcher's interest and provides useful information. Some examples of information that are typically identified, mined, or extracted from biomedical literature include an entity interacting with an entity like a drug interacting with a drug or a protein interacting with a protein. It also includes a relationship between an entity with another entity like between a protein and a residue or a gene or any temporal relationship in addition to any other bio-entities that are part of the event. Hence, this process automates and simplifies the task of offering only useful textual data required to the researcher by eliminating cumbersome manual searching efforts [14,15]. These mined events from vast literature are important and have many real-world applications like database

curation, constructing ontology, semantic web search, interactive systems etc.

Named entity recognition (NER), is a subset of the event extraction task that involves identifying and detection of references to entities like genes, and proteins [16]. It is also a research area of interest which has gained a lot of traction over the last decade.

This is because there is still a large gap of >10% in the F1- score using the best ML algorithms for biomedical NER vs. those used for any general-purpose NER. Hence, researchers are exploring various methods to narrow the gap using better pre-processing and feature extraction techniques. Most approaches for effectively identifying named entities (NEs) in biomedical literature fall under three categories, namely heuristic rule-based, dictionary-based, and statistical machine learning-based approaches. But the results to date have been far from satisfactory which suggests that there is still no robust, generalized implementation of any NER system nor any algorithm which can be singled out that can provide higher performance.

This paper is divided into the following sections: Section 2 explains a biomolecular event extraction task, and Section 3 discusses and compares the performance of the existing systems. Section 4 discusses new architectures and suggests two novel approaches for developing a robust BioNLP model. We conclude with a note summarizing the approaches, existing challenges, and future research directions in Section 5.

2. BIOMOLECULAR EVENT EXTRACTION

Identifying and isolating semantic relations is the primary task when it comes to biomedical text mining. Here, information is extracted from a vast volume of document sets or big data like scientific literature or patient records [17-19]. The information contains, apart from other things, statements of interactions between NEs, like the effect of drugs on a patient, cellular protein movements etc. Relation extraction and event extraction are the two ways for getting such data. Relations can be typed, directed, or pairwise links type between defined named entities.

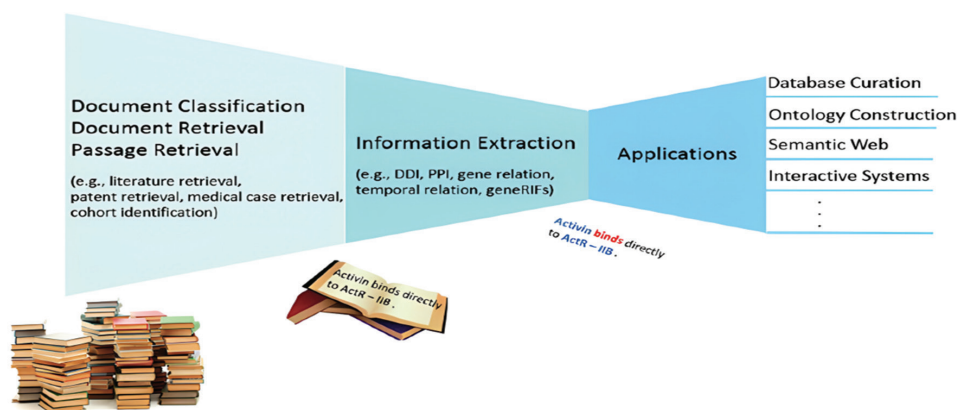


Fig. 2. NLP-based data-channeling process and applications [20]

Event extraction is another form of relation extraction in which events can combine two or more entities, they have a trigger word which is usually a verb and they also sometimes act as arguments for other events [21]. Though events are efficient in capturing the semantics of text more precisely extracting them is an extremely complicated task. Table 1 explains the event extraction process. Post-pre-processing and feature extraction, first, the NEs (e.g., TGF-beta) are identified. The next stage is to detect trigger words and labels via annotation of the phrases and finally the process reconstructs the events if the edges are clear or not overlapping.

A typical biomedical event extraction task flow-chart is shown in (Fig. 3). The first two steps are pre-processing and feature extraction followed by named entity recognition [22]. Trigger detection (which identifies event triggers and types) and edge detection (links event triggers with arguments) are the two sub-steps of the main event detection step. Some researchers combine these two steps to reduce cascading errors which helps in improving performance. Post-processing is the final stage that helps in refining the final event structure outcome [23,24].

All the steps are explained in the following sections outlining various tools and approaches used.

Table 1. A biomolecular event extraction task workflow

Phrase: "TGF-beta mediates RUNX induction and FOXP3 is efficiently up-regulated by RUNX1 and RUNX3"		
Entities: "TGF-beta, RUNX, FOXP3, RUNX1 and RUNX3 (proteins)"		
Step 1	Entity recognition	Entities: TGF-beta, RUNX, FOXP3, RUNX1 and RUNX3 (proteins)
Step 2	Trigger detection	Trigger 1: induction (gene expression)
		Trigger 2: mediates (positive regulation)
		Trigger 3: up-regulated (positive regulation)
Step 3	Edge detection	Edge 1: induction (gene expression); theme: RUNX(protein)
		Edge 2: mediates (positive regulation); cause: RUNX (protein); theme: above gene expression event
		Edge 3: up-regulated (positive regulation); cause: RUNX1 and RUNX3; theme: FOXP3 (protein)
Step 4	Reconstruct event	Event 1: induction (gene expression); theme: RUNX (protein)
		Event 2: mediates (positive regulation); cause: RUNX (protein); theme: above gene expression event
		Event 3: up-regulated (positive regulation); cause: RUNX1; theme: FOXP3 (protein)
		Event 4: up-regulated (positive regulation); cause: RUNX3; theme: FOXP3 (protein)

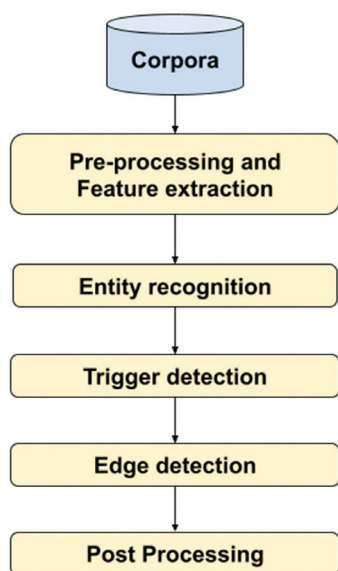


Fig. 3. Flow-chart of a typical biomolecular event extraction task

2.1. CORPORA

There are many corpora available for biomolecular event extraction tasks. Some of the most popular ones used are enumerated below:

- GENIA Event dataset [25]: This is made available by the BioNLP shared task [26] organizers openly

to all researchers. This consists of human-curated complex event events. It has 1000 Medline paper abstracts, which in turn have 9372 sentences and 36114 events have been identified from it.

- BioInfer Dataset [27]: It is a publicly available dataset consisting of manually annotated corpus and other resources for extraction of information. It has 1,100 sentences from biomedical research abstracts.
- PPI dataset [28]: A Protein-protein interaction corpus complements available training data and is not as elaborate as the event corpora. LLL, AIMed and BioCreative are the most relevant PPI corpora used.

2.2. PRE-PROCESSING AND FEATURE EXTRACTION

Datasets are getting more complex by the day and it may be required to work with datasets containing hundreds of features. If the number of features equals or becomes more than the number of observations stored in a dataset, it could lead to overfitting. Hence pre-processing and applying feature extraction techniques are necessary. Pre-processing, a mandatory step in any text mining pipeline involves reading the data from its original format to an internal representation. It involves a set of common NLP tasks, from sentence segmentation and tokenization to part-of-speech tag-

ging, chunking, and linguistic parsing. In addition to these, the biomolecular event extraction task also involves removing co-references, sentence simplification etc. to improve accuracy [29,30]. The most commonly used pre-processing frameworks are NLTK (<http://www.nltk.org/>), Stanford CoreNLP (<http://nlp.stanford.edu/software/corenlp.shtml>) and Apache OpenNLP (<https://opennlp.apache.org/>).

Table 2. Commonly used features in the event detection phase

Feature Groups	Features	Trigger recognition	Edge detection
Token-based	Parts-of-speech (POS)	Yes	Yes
	Lemma	Yes	No
	Orthographic	Yes	No
	n-grams	Yes	No
	Word shape	Yes	No
	Prefixes/suffixes	Yes	No
Contextual features	Number of entities	Yes	No
	BoW counts	Yes	No
	Windows or conjunctions of features	Yes	No
Dependency-based features	Number and type of dependency edges	Yes	No
	Words, lemmas, and POS tags in the dependency path	Yes	Yes
	N-grams in the dependency path	Yes	Yes
External features	WordNet lemmas	No	No
	Trigger lexicon	No	No
	Entity lexicon	No	No

Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones, which are then discarded. Feature extraction techniques provide below advantages:

- Accuracy improvement [31-34]
- Overfitting risk reduction
- Speed up in training
- Improved data visualization
- Increase in explainability of a model

There are many feature extraction techniques used. The commonly used features in the main event detection stage are shared in Table 2. They can be divided into four feature groups namely token-based, contextual, dependency and external resources which are explained below.

- Token-based features: They capture a specific feature for every token like part-of-speech, lemma, prefix, suffix etc. In part-of-speech, each word in a sentence is tagged with its POS like noun, verb, adjective etc. However, its main drawback is ambiguity. In the biomedical domain, many frequently used words have multiple meanings hence multiple POS. Lemmatization breaks a word down to its root meaning to identify similarities. But it is a slow and

time-consuming process as it involves performing morphological analysis and deriving the meaning of words from a dictionary. Orthographic rules are general rules used when breaking a word into its stem and modifiers. However, many languages have various levels of orthographic depth and orthographies that are highly irregular, and difficult, and where sounds cannot be predicted from the spelling. n-grams refer to a sequence of N words or characters. More is not necessarily better as in some cases, having too many features will result in a less optimal model. Base forms of the input words, word prefixes and suffix features, and basic word form/shape features have also been used for trigger recognition with varying degrees of success.

- Contextual features: They offer knowledge about the sentence or its neighborhood where the token exists. The same word or phrase can also have different meanings according to the context of a sentence or many words. The number of tokens in the sentence, no. of entities or the bag-of-words count are some of the contextual features extracted from sentences. A bag of words is a representation of text that describes the occurrence of words within a document. It just keeps track of word counts and disregards the grammatical details and the word order. It is called a 'bag' of words because any information about the order or structure of words in the document is discarded. The model is only concerned with whether known words occur in the document, not where in the document. A drawback of the BoW model is that it ignores the location information of the word. Also, the model does not respect the semantics of the word. The range of vocabulary is a big issue faced by the bag-of-words model.
- Dependency-based features: They provide the Grammar relation between two words and can be extracted from a dependency relationship graph of a sentence. Commonly used features include the number and type of dependency edges between two tokens and the words, lemmas, POS and n-gram characteristics in the dependency path. The drawbacks of dependency parsing are first, semantic actions cannot be performed while making a prediction. The actions must be delayed until the prediction is known to be a part of a successful parse. Secondly, precise error reporting is not possible. A mismatch merely triggers backtracking.
- External features: Here external knowledge is augmented from sources like WordNet, lexicons both trigger and entity from the dictionary or other sources. This helps to improve accuracy in cases where the no. of words is long, has many meanings etc. In linguistics, the canonical form or morphological form of a word is called a lemma. To find a synonym as well as an antonym of a word, one can look up lemmas in WordNet. The drawbacks are that the vocabulary it contains is broad and thus

ambiguous. Lexicons are collections of domain-specific key phrases (also known as entities) that can be attached to a flow. A Lexicon can be seen as a dictionary, which allows the virtual agent to 'understand' specific words. The drawback of a lexicon-based feature extraction model is scalability.

2.3. NAMED ENTITY RECOGNITION

NER detects references to entities like genes, proteins, and chemical compounds [35,36]. It is a challenging task in the biomedical domain as new entities keep getting added and there is a lack of a complete dictionary. Also, it may so happen that a word may refer to two different entities based on their context. It is also seen that many biological NEs have different spelling forms, have abbreviations or are exceptionally long which come in the way of classifying them or identifying their boundaries correctly. They may also be nested in other entities, which require more effort to identify such NEs. There are a few NER systems proposed, but the best results are in the 85% F1-score range [37]. Some of the most popular NER tools used are summarized in Table 3.

Table 3. Most popular NER tools used

Gene/ Proteins	Chemical compounds	Gene protein and disorders
Gimli	SCAI	BioEnEx
NERSuite	ChemSpot	BANNER
AllAGMT	Neji	-
GNAT	-	-
GeNo	-	-
ABNER	-	-

2.4. TRIGGER DETECTION

A lot of research is focused on this area as the subsequent steps' effectiveness is based on the information outcome from this step. It involves identification of event triggers and their types. As seen from Table 1, the trigger word 'induction' defines an event of type gene expression and the trigger word 'mediates' defines an event of type Positive Regulation. Complexity arises when sentences may contain many related events, Negative Regulation and the same trigger could also indicate diverse types of events, based on the context [38-40]. Table 4 summarizes the different methodologies adopted for Trigger detection which are broadly classified as rule-based based, dictionary-based and machine learning based. It is observed that rule-based methods return low recall rates as defining detailed rules requires a lot of effort and some of the rules are too hard to accommodate semantic paraphrases. Dictionary-based approaches contain a dictionary with trigger words and their corresponding classes (event types) to identify and enumerate event triggers. Researchers have also used a hybrid approach by combining rule-based and dictionary-based approaches [41,42].

Table 4. Different methodologies adopted for trigger detection

Rule-based	Dictionary-based	Machine Learning based				References
		SVM	CRF	VSM	MEMM	
√	√	×	×	×	×	[51]
×	√	×	√	×	×	[8]
×	×	√	×	×	×	[52]
×	√	×	×	×	×	[53]
√	√	×	×	×	×	[41]
√	√	√	×	×	×	[54]
√	×	×	×	×	×	[55]
√	×	√	√	×	×	[56]
×	×	√	×	√	×	[57]
×	×	×	√	×	√	[58]
×	×	×	√	×	×	[59]
×	×	×	√	×	×	[60]
×	×	×	√	×	×	[61]
×	×	×	√	×	×	[62]
×	×	×	√	×	×	[63]
×	×	×	√	×	×	[64]
×	×	×	√	×	×	[65]
×	×	×	×	√	×	[66]

Machine Learning based approaches like Support vector machine (SVM) and its variant kernels like linear, radial basis function, polynomial and convolution; Conditional random field (CRF); Value stream matching (VSM) and Maximum entropy Markov model (MEMM) have been most successful for trigger detection and offer higher recall rates [43-46].

2.5. EDGE DETECTION

It predicts arguments in an event which can be named entities, or another event represented as a different trigger word. It is also known as event theme construction. In Table 1, for the sentence, there are three edges identified. Again, rule, dictionary and machine learning-based methods have been suggested to address this task. Like trigger detection, there have been more efforts in using ML algorithms by treating the edge detection problem as a supervised multi-class classification problem. Also, many studies are based on hybrid approaches by using a combination of the above methods or using ensemble methodology [47,48] as shown in Table 5.

Table 5. Hybrid approaches used for trigger detection

Ensemble	Base Learners	CI-Optimized algorithms	F1-score (%)	References
Yes	1	Yes	55.64	[67]
Yes	3	Yes	57.58	[47]
Yes	5	No	66.34	[68]

Here, the thought process is not to choose the ‘best’ classifier always, as it may not be representative of all data. In this, a few classifiers are trained in a standalone mode. The output of all the classifiers is combined into an ensemble and the result is chosen according to some criteria. A problem arises on how to choose the best one, as more than one classifier may also meet the criteria of similar training accuracy. Methods of the combination include bagging, boosting, voting, stacked generalization, and cascading. This leads to improvement in prediction accuracy [49].

Researchers have also used optimization techniques to choose the right weights for voting using various nature-inspired Computational Intelligence (CI) algorithms [50]. Table 6 summarizes the various approaches used for Edge detection.

Table 6. Various approaches used for Edge detection

Rule-based	Dictionary-based	Machine Learning based			References
		SVM	CRF	HVS	
√	√	x	x	x	[54]
√	x	√	x	x	[60]
√	x	x	x	x	[8]
x	x	√	x	x	[53]
√	x	x	x	x	[41]
√	x	x	x	√	[51]
x	x	√	x	x	[58]
x	x	x	√	x	[8]
x	x	√	x	x	[59]
x	x	√	x	x	[60]
x	x	√	x	x	[62]
x	x	√	x	x	[63]
x	x	√	x	x	[69]

3. EXISTING SYSTEM

Many bio-text mining campaigns are running successfully for years. Some of the most popular ones are listed below. They address many aspects from NER to the biological phenomenon to text categorization

- KDDCup
- TREC-Genomics
- JNLPBA for NER
- BioCreative for extraction of NER, PPI, text categorization
- BioNLP Shared tasks

We discuss and compare the existing state-of-the-art systems and the approaches used in this section. For a similar comparison, the results achieved by various systems for event extraction on the BioNLP shared tasks are compared [70,71]. The BioNLP-ST uses the Genia event (GE) corpus with sub-tasks. It extracts events from both complete papers and abstracts. The sub-tasks are divided into three categories, (i) Core event

extraction (GE), (ii) Event enrichment (GE 2) and (iii) Negation/Speculation detection (GE 3). Table 7 lists the shared tasks released by the committee in 2013 [72-74].

Table 8. Results obtained in the BioNLP-ST GE task by top systems (in %age)

System, Year, Reference	Simple	Event type binding	Regulation	Total
Univ. of Turku, 2009 [52]	70.21	44.41	40.11	52
Miwa, 2010 [53]	70.44	52.62	40.6	53.3
FAUST, 2011 [75]	72.85	51.05	46.97	57.5
EVEX [63]	76.59	42.88	38.41	51
TEES 2.1, 2013 [76]	76.82	43.32	38.05	50.7
BioSM [77]	76.11	49.76	35.8	50.7

Table 9. Results from BioNLP-ST GE 2 task by top systems (in %age)

System, Year, Reference	Site	Localization	Total
Univ. of Turku, 2009 [52]	71.43	36.59	44.5
FAUST, 2011 [75]	50	50	52.8
EVEX, 2013 [63]	50	50	31.2
TEES 2.1 [76]	50	50	32.5

Table 10. Results obtained in the BioNLP-ST GE 3 task by top systems (in %age)

System, Year, Reference	Document	Negation	Speculation	Total
Concord, 2009 [54]	Abs	23.13	25.27	24.17
Univ. of Turku, 2011 [55, 78]	Abs	30.4	25.64	28.08
EVEX, 2013 [63]	FP	27.04	23.92	25.22
TEES 2.1 [77]	FP	27.3	23.61	25.15

Table 11. Benchmark performance of training GloVe on CPU vs. GPU

Architecture	Specifications	Performance	Speed-up
CPU	i7-6800K, 8-cores	13.56 min/epoch	-
GPU	NVidia GTX 1070	1.22 min/epoch	11X

4. NEW ARCHITECTURES

Implementing CNNs and RNNs, though help in improving accuracy but is impacted on account of training time, computational cost and memory requirements. BERT (Bidirectional encoder representations from Transformers) which is an encoder stack of the

Transformer architecture has emerged as a strong contender for BioNLP tasks as it is pre-trained on a large but generic corpus. However, to make it suitable for BioNLP, it needs to be suitably trained on additional Biomedical corpus which is time-consuming. Hence, newer hardware architectures are required as an alternative to standard x86 processors. High-performance accelerators exist for many tasks today, that can be explored for NLP too. Current models focus only on accuracy and seldom on the issues concerning the above three impact criteria mentioned. Nvidia Graphical processing unit (GPU), Intel Xilinx Field programmable gate array (FPGA) and Google’s Tensor Processing Units (TPU) are some of the accelerator options that provide high-performance computing (HPC) at a fraction of the cost. We discuss the benefits of one such accelerator in this paper, the Nvidia GPU, compare it with a standard x86 processor and explain briefly how it can accelerate complex BioNLP tasks.

4.1. GPU ACCELERATED BIONLP

Many tasks in NLP can exploit the massive parallelism offered by the GPU. GPUs were initially used only for graphics or visualization tasks. Due to the massive number of lightweight cores (SMs) as shown in (Fig. 4), in early 2000 Nvidia introduced the CUDA parallel programming environment, which revolutionized high-performance computing. CUDA is C-like programming which allowed researchers to port their compute-hungry codes onto the GPU. Single-instruction multiple-thread (SIMT) execution model emerged and with the introduction of shared memory and a heterogeneous CPU-GPU architecture, all heavy-duty compute or mathematical tasks were off-loaded to the GPUs for processing thereby accelerating the entire code.

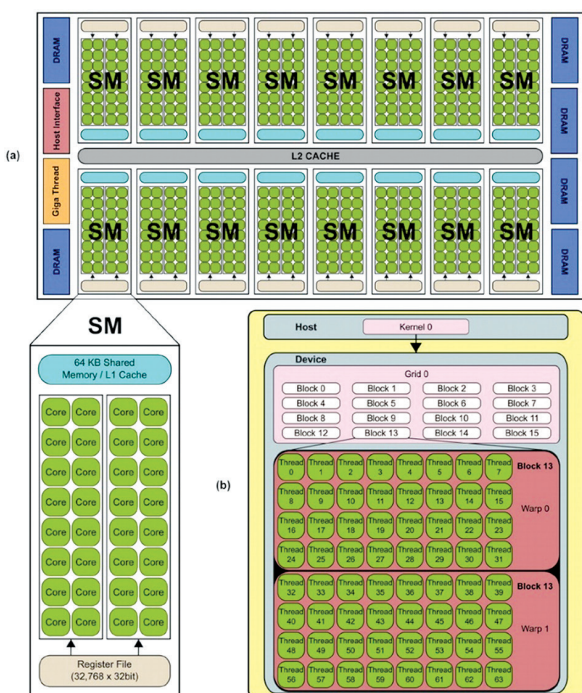


Fig. 4. The Nvidia GPU architecture [79]

Many GPU-accelerated NLP toolkits exist including GloVe (Global vectors for word representation), which is an extremely popular unsupervised learning algorithm for words. Especially for NLP, once the text is hashed, GPUs can offer accelerated results from the voluminous scientific literature much faster (large no. of words per millisecond) as compared to an x86 CPU. (Fig. 5) shows the difference between a CPU which has lesser ALUs compared to a GPU which explains this significant difference in computational power.

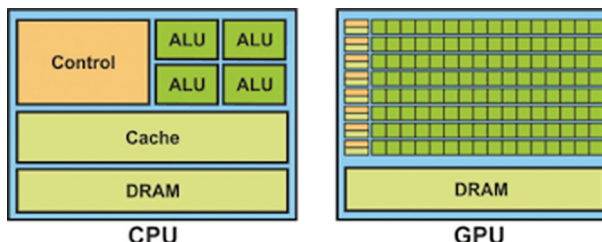


Fig. 5. Differences between a CPU and GPU [79]

The goal of a CPU is different from a GPU in that a CPU must be capable of processing everything, serial or parallel but the GPUs are ground-up built for parallel tasks only, and NLP falls in this category. One of the drawbacks of using GPUs is the precision overhead. Precision is particularly important in operations like NLP, computing gradients etc. It is found that GPU speeds with double precision are 2-32x slower. Also, it consumes a lot of memory and power. The solution for this lies in using mixed precision for NLP tasks and lower-power gaming cards like the GTX series. Table 11 shows the benchmark performance of training GloVe on a CPU vs. a CPU+GPU system.

Very few implementations of CNN and RNN models have been tried on accelerators, which can help achieve up to 10 Tera operations (TOPs) per second on such semiconductor chips [80]. Each of these chips has its unique hardware architecture and programming environments to launch thousands of threads to parallelize the applications. But we still need large-scale HPC clusters with large numbers of nodes in a datacenter to achieve an inference efficiency of millions of words per second. Authors in [81] implemented a Dynamic multi-pooling convolutional neural network (DMCNN) that took 1.0 GOPs for processing 30 words in a single sentence. Normal CPUs are unable to keep pace with RNN computations on account of irregular computation and memory access. These accelerators though desired have some limitations as they are rated at higher power and provide lower performance per watt as the complexity of neural networks increases. Thus, achieving a balance of accuracy, computational cost and memory usage needs to be targeted and this is an area where not much research has been conducted.

4.1. RECOMMENDED ARCHITECTURES

Based on an extensive literature study and results of best-performing models, an ensemble model of ML classifiers and a domain-specific BERT pre-trained us-

ing GPU for reducing the training time on additional biomedical corpora (like PubMed and PubMed Central extracts) are the two key architectures recommended for NER, the most challenging sub-task in Biomolecular

event extraction. We propose high-level reference architecture and methodology of an ensemble classification model and NeRBERT- a biomedical NER tagger in (Fig. 6) and (Fig. 7) respectively.

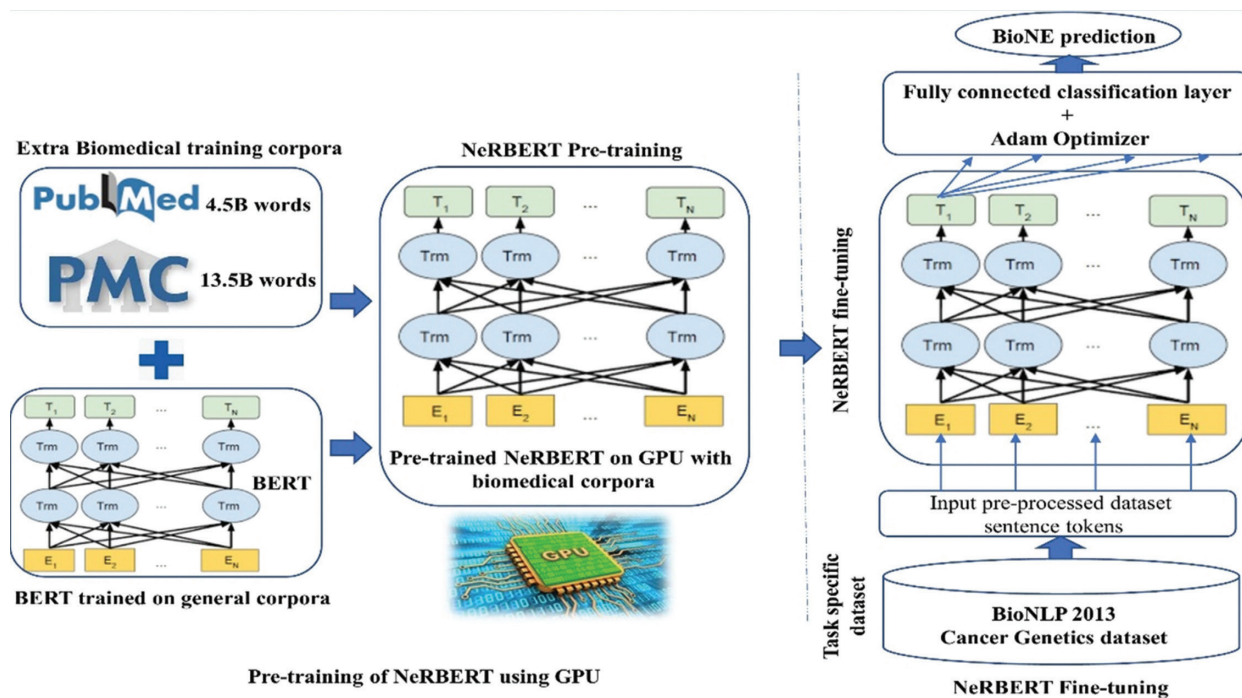


Fig. 7. Recommended architecture of a pre-trained and fine-tuned NeRBERT model

5. CONCLUSION

The paper provides a systematic summarization of the latest research in the field of biomolecular event extraction. It discusses the various techniques used by researchers due to the challenging nature of the task, ambiguity, and heterogeneity in biomedical literature. The results summarized demonstrate that this is still a challenging proposition, despite the slow and steady improvements. Although performance results of 80% in the F1-score were obtained in the identification of simpler events, there is still minimal extraction of more complex events such as binding and regulatory events. Although major efforts have been made to recognize the events, the best performance achieved remains 30–40% lower than that for simple events. Many techniques have been proposed, which include simple parsing, pattern-matching, machine learning and deep learning methods. Researchers have adopted multi-stage approaches wherein the second stage fine-tunes the output from the first stage either by using some rules or techniques like ensemble classification which is one of the architectures recommended. Alternatively, BERT has also emerged as an architecture with great promise. We propose two reference architectures based on these in terms of accuracy, computational cost, and memory usage for developing a robust GPU-accelerated BioNLP model for biomolecular event extraction. The BERT model should be domain-specific and trained on additional biomedical corpora using a

GPU to reduce training time. However, key challenges to mitigate still exist which involve extracting complex regulatory events, resolving cross-references, and defining negation and speculation. Also, machine learning approaches like transfer learning and other newer hardware architecture like FPGA and TPU have not been employed much which provide a lot of scopes to accelerate the event extraction pipeline. Despite these challenges, available research can still help in curating pipelines using text-mined data, constructing networks, ontologies, and knowledge bases.

6. REFERENCES:

- [1] M. S. Simpson, D. Demner-Fushman, "Biomedical text mining: a survey of recent progress", *Proceedings of Mining Text Data*, Springer, 2012, pp. 465-517.
- [2] C. Li, M. Liakata, D. Rebholz-Schuhmann, "Biological network extraction from scientific literature: state of the art and challenges", *Briefings in Bioinformatics*, Vol. 15, No. 5, 2014, pp. 856-877.
- [3] A. Manconi, E. Vargiu, G. Armano, L. Milanese, "Literature retrieval and mining in bioinformatics: state of the art and challenges", *Advances in Bioinformatics*, 2012, pp. 10-10.

- [4] S. Ananiadou, P. Thompson, R. Nawaz, "Event-based text mining for biology and functional genomics", *Briefings in Functional Genomics*, Vol. 14, No. 3, 2015, pp. 213-230.
- [5] L. Hirschman, G. A. P. C. Burns, M. Krallinger, "Text mining for the Biocuration workflow", *Database: The Journal of Biological Databases and Curation*, 2012.
- [6] D. Campos, S. Matos, J. L. Oliveira, "Current methodologies for biomedical named entity recognition", John Wiley & Sons, 2013, pp. 839-868.
- [7] C. N. Arighi, Z. Lu, M. Krallinger, "Overview of the Biocreative III workshop", *BMC Bioinformatics*, Vol. 12, 2011, Supplement 8, S1.
- [8] D. Mackinlay, D. Martinez, T. Baldwin, "Biomedical event annotation with CRFs and precision grammars", *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, June 2009, pp. 77-85.
- [9] E. Ghosh, H. Naja, H. Abdulrab, M. Khalil, "Ontology Learning Process as a Bottom-up Strategy for Building Domain-specific Ontology from Legal Texts", *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*, 2017, pp. 473-480.
- [10] I. Segura-Bedmar, P. Martínez, M. Herrero-Zazo, "SemEval-2013 Task 9: Extraction of Drug-Drug Interactions from Biomedical Texts", *Proceedings of the Seventh International Workshop on Semantic Evaluation*, Atlanta, GA, USA, June 2013, pp 341-350.
- [11] S. Ananiadou, S. Pyysalo, J. Tsujii, D. B. Kell, "Event extraction for systems biology by text mining the literature", *Trends in Biotechnology*, Vol. 28, No. 7, 2010, pp. 381-390.
- [12] A. Prodromidis, P. K. Chan, P. S. J. Stolfo, "Meta-learning in distributed data mining systems: Issues and approaches", *Advances in distributed and data mining*, AAAI Press, 2016.
- [13] P. Thompson, S. A. Iqbal, J. Mcnaught, S. Ananiadou, "Construction of an annotated corpus to support biomedical information extraction", *BMC Bioinformatics*, Vol. 10, 2009, pp. 349-349.
- [14] D. Zhou, S. Jian, "Deep Knowledge Resources in Biomedical Name Recognition", *Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications*, Geneva, Switzerland, 2014.
- [15] H. G. Lee, H. C. Cho, M. J. Kim, J. Y. Lee, G. Hong, H. C. Rim, "A multi-phase approach to biomedical event extraction", *Proceedings of the Workshop on BioNLP*, Boulder, CO, USA, June 2009, pp. 107-110.
- [16] S. Burr, "Biomedical Named Entity Recognition Using Conditional Random Fields and Novel Feature Sets", *Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications*, Geneva, Switzerland, 2004.
- [17] S. Kulick, A. Bies, M. Liberman, "Integrated Annotation for Biomedical Information Extraction", *Proceedings of HLT/NAACL 2014 Biobank Workshop*, 2014, pp. 61-68.
- [18] D. Mcclosky, M. Surdeanu, C. Manning, "Event extraction as dependency parsing for BioNLP", *Proceedings of the BioNLP Shared Task Workshop*, Portland, OR, USA, June 2011, pp. 41-45.
- [19] M. Miwa, S. Pyysalo, T. Ohta, S. Ananiadou, "Wide coverage biomedical event extraction using multiple partially overlapping corpora", *BMC Bioinformatics*, Vol. 14, No. 1, 2013, pp. 175-175.
- [20] C. C. Huang, Z. Lu, "Community challenges in biomedical text mining over 10 years: success, failure, and the future", *Briefings in Bioinformatics*, Vol. 17, No. 1, 2016, pp. 132-144.
- [21] S. Riedela, D. Mcclosky, M. Surdeanu, A. Mccalluma, C. Manning, "Model combination for event extraction in BioNLP", *Proceedings of BioNLP Shared Task 2011 Workshop*, Portland, OR, USA, June 2011, pp. 51-55.
- [22] J.-D. Kim, T. Ohta, Y. Tsuruoka, "Introduction to the Bio- Entity Recognition Task at JNLPBA", *Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications*, Geneva, Switzerland, August 2014, pp. 70-75.
- [23] D. Mcclosky, "Any domain parsing: automatic domain adaptation for natural language parsing", Providence, RI, USA. AAI3430199, 2010.
- [24] E. Buyko, E. Beisswanger, U. Hahn, "The genereg corpus for gene expression regulation events- an

- overview of the corpus and its in-domain and out-of-domain interoperability”, Proceedings of the 7th International Conference on Language Resources and Evaluation, Valletta, Malta, May 2010, pp. 1921-1921.
- [25] J. D. Kim, T. Ohta, K. Oda, J. I. Tsujii, “From text to pathway: corpus annotation for knowledge acquisition from biomedical literature”, Proceedings of the Asia-Pacific Bioinformatics Conference, Imperial College Press, 2008, pp. 165-176.
- [26] J. Bjerne, T. Salakoski, “Generalizing biomedical event extraction”, Proceedings of the BioNLP Shared Task 2011 Workshop, Portland, OR, USA, June 2011, pp. 183-191.
- [27] S. Pyysalo, F. Ginter, J. Heimonen, “BioInfer: a corpus for information extraction in the biomedical domain”, BMC Bioinformatics, Vol. 8, 2007, pp. 50-50.
- [28] The LLL corpus, 2015.
- [29] D. Tikk, P. Thomas, P. Palaga, J. Hakenberg, U. Leser, “A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature”, PLoS Computational Biology, 2010.
- [30] A. Berger, S. Della Pietra, V. Della Pietra, “A maximum entropy approach to natural language processing”, Computational Linguistics, Vol. 22, No. 1, 2016, pp. 39-71.
- [31] Y. Tsuruoka, T. Yuka, J.-D. Kim, T. Ohta, J. McNaught, S. Ananiadou, J. Tsuji, “Developing a Robust Part-of-Speech Tagger for Biomedical Text”, Proceedings of the 10th Panhellenic Conference on Informatics, 2015, pp. 382-392.
- [32] T. Zhang, F. Damerou, D. Johnson, “Text chunking based on a generalization of Winnow”, Journal of Machine Learning Research, 2012, pp. 615-637.
- [33] J.-D. Kim, Y. Wang, N. Colic, S. H. Beak, Y. H. Kim, M. Song, “Refactoring the Genia event extraction shared task toward a general framework for IE-driven KB development”, Proceedings of the Fourth BioNLP Shared Task Workshop, Berlin, Germany, August 2016, pp. 23-31.
- [34] R. McDonald, F. Pereira, S. Kulick, S. Winters, Y. Jin, P. White, “Simple algorithms for complex relation extraction with applications to biomedical”, Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA, June 2005.
- [35] G. Li, K. E. Ross, C. N. Arighi, Y. Peng, C. H. Wu, K. V. Shanker, “MirTex: A text mining system for MiRNA-Gene relation extraction”, PLoS Computational Biology, Vol. 11, No. 9, 2015, pp. 104-108.
- [36] K. Yoshikawa, S. Riedel, T. Hirao, Asahara, Y. Matsumoto, “Coreference based event-argument relation extraction on biomedical text”, Proceedings of the Fourth International Symposium on Semantic Mining in Biomedicine, 2010.
- [37] D. Campos, S. Matos, J. L. Oliveira, “Biomedical named entity recognition: a survey of machine-learning tools”, Theory and Applications for Advanced Text Mining, 2012, pp. 175-195.
- [38] J. Berant, V. Srikumar, P. C. Chen, “Modeling biological processes for reading comprehension”, Proceedings of the Empirical Methods in Natural Language Processing, Doha, Qatar, October 2014.
- [39] L. Hirschman, M. Krallinger, A. Valencia, “Chendner: The drugs and chemical names extraction challenge”, Journal of Cheminformatics, Vol. 7, 2015, S1.
- [40] J. D. Nguyen, M. Kim, T. Miwa, J. Matsuzaki, J. Tsujii, “Improving protein coreference resolution by simple semantic classification”, BMC Bioinformatics, Vol. 13, 2012, pp. 304-304.
- [41] Q. Leminh, S.N. Truong, Q.H. Bao, “A pattern approach for biomedical event annotation”, Proceedings of the BioNLP Shared Task 2011 Workshop, Association for Computational Linguistics, Portland, OR, USA, June 2011, pp. 149-150.
- [42] J. Gama, P. Brazdil, “Cascade generalization”, Machine learning, Vol. 41, No. 3, 2014, pp. 315-343.
- [43] A. Ozgur, D. Radev, “Supervised classification for extracting biomedical events”, Proceedings of the Workshop on BioNLP, Boulder, CO, USA, June 2009, pp. 111-114.
- [44] J. Lafferty, A. McCallum, F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labelling sequence data”, Proceedings of the Eighteenth International Conference on Machine Learning, June 2001, pp. 282-289.

- [45] A. Airola, S. Pyysalo, J. Bjorne, T. Pahikkala, F. Ginter, T. Salakoski, "All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning", *BMC Bioinformatics*, Vol. 9, 2008, S2.
- [46] D. Tikk, P. Thomas, P. Palaga, J. Hakenberg, U. Leser, "A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature", *PLoS Computational Biology*, 2010.
- [47] A. Majumder, A. Ekbal, S.K. Naskar, "Feature selection and class-weight tuning using genetic algorithm for bio-molecular event extraction", *Proceedings of the NLDB*, 2017.
- [48] A. Majumder, A. Ekbal, S.K. Naskar, "Biomolecular event extraction using a stacked generalization-based classifier", *Proceedings of the 13th International Conference on Natural Language Processing*, 2016, pp. 55-64.
- [49] K. Ting, I. Witten, "Issues in stacked generalization", *Journal of Artificial Intelligence Research*, Vol. 10, 2013, pp. 271-289.
- [50] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, "A fast and elitist multi objective genetic algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, 2002, pp. 182-197.
- [51] H. Kilicoglu, S. Bergler, "Effective bio-event extraction using trigger words and syntactic dependencies", *Computational Intelligence*, Vol. 27, No. 4, 2011, pp. 583-609.
- [52] J. Bjorne, F. Heimonen, Ginter, "Extracting complex biological events with rich graph-based feature sets", *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, Boulder, CO, USA, June 2009, pp. 10-18.
- [53] M. Miwa, R. Saetre, J.D. Kim, J. Tsujii, "Event extraction with complex event classification using rich features", *Journal of Bioinformatics and Computational Biology*, Vol. 8, No. 1, 2010, pp. 131-146.
- [54] H. Kilicoglu, S. Bergler, "Syntactic dependency-based heuristics for biological event extraction", *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, Association for Computational Linguistics, Boulder, CO, USA, June 2009, pp. 119-127.
- [55] A. D. Casillas, K. D. Illaraza, M. Gojenola, G. Oronoz, Rigau, "Using kybots for extracting events in biomedical texts", *Proceedings of the BioNLP Shared Task 2011 Workshop*, Portland, OR, USA, June 2011, pp. 138-142.
- [56] S. V. Landeghem, B. De, Y. Baets, Y. D. Peer, Saeys, "High-precision bio-molecular event extraction from text using parallel binary classifiers", *Computational Intelligence*, Vol. 27, No. 4, 2011, pp. 645-664.
- [57] D. Martinez, T. Baldwin, "Word sense disambiguation for event trigger word detection in biomedicine", *BMC Bioinformatics*, Vol. 12, 2011.
- [58] D. Zhou, Y. He, "Biomedical events extraction using the hidden vector state model," *Artificial Intelligence in Medicine*, Vol. 53, No. 3, 2011, pp. 205-213.
- [59] M. Miwa, P. Thompson, S. Ananiadou, "Boosting automatic event extraction from the literature using domain adaptation and coreference resolution", *Bioinformatics*, Vol. 28, No. 13, 2012, pp. 1759-1765.
- [60] F. Bjorne, T. Ginter, Salakoski, "University of Turku in the BioNLP'11 Shared Task", *BMC Bioinformatics*, Vol. 13, 2012, pp. 13.
- [61] L. Qian, G. Zhou, "Tree kernel-based protein-protein interaction extraction from biomedical literature", *Journal of Biomedical Informatics*, Vol. 45, No. 3, 2012, pp. 535-543.
- [62] J. Wang, Q. Xu, H. Lin, Z. Yang, Y. Li, "Semi-supervised method for biomedical event extraction", *Proteome Science*, Vol. 11, 2013, p. 17.
- [63] K. Hakala, S.V. Landeghem, T. Salakoski, "EVEX: application of a large-scale text mining resource to event extraction and network construction", *Proceedings of the BioNLP Shared Task 2013 Workshop*, Association for Computational Linguistics, Sofia, Bulgaria, August 2013, pp. 26-34
- [64] Y. Zhang, H. Lin, Z. Yang, J. Wang, Y. Li, "Biomolecular event trigger detection using neighborhood hash features", *Journal of Theoretical Biology*, Vol. 318, 2013, pp. 22-28.
- [65] X. Liu, A. Bordes, Y. Grandvalet, "Biomedical event extraction by multi-class classification of pairs of

- text entities”, Proceedings of the BioNLP Shared Task 2013 Workshop, Association for Computational Linguistics, Sofia, Bulgaria, August 2013, pp. 45-49.
- [66] D. Campos, Q. C. Bui, S. Matos, J. L. Oliveira, “Tri-gNER: automatically optimized biomedical event trigger recognition on scientific documents”, Source Code for Biology and Medicine, Vol. 9, No. 1, 2014.
- [67] L. Li, Y. Wang, D. Huang, “Improving feature-based biomedical event extraction system by integrating argument information”, Proceedings of the BioNLP Shared Task 2013 Workshop, Sofia, Bulgaria, August 2013, pp. 109-115.
- [68] M. Bali, P. V. R Murthy, “Bio-Molecular Event Extraction Using Classifier Ensemble-of-Ensemble Technique”, Data Management, Analytics and Innovation, Springer, 2021, pp.445-462.
- [69] J. Xia, A. C. Fang, X. Zhang, “A novel feature selection strategy for enhanced biomedical event extraction using the Turku system”, BioMed Research International, 2014.
- [70] L. Hunter, Z. Lu, J. Firby, W. A. Baumgartner, H. L. Johnson, P. Ogren, V. K Cohen, “Opendmap: an opensource, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression”, BMC Bioinformatics, Vol. 9, 2008, pp. 78-78.
- [71] K. Yoshikawa, S. Riedel, T. Hira, “Coreference based event-argument relation extraction on biomedical text”, Journal of Biomedical Semantics, Vol. 2, 2011, S6.
- [72] M. Krallinger, F. Leitner, C. Rodriguez-Penagos, A. Valencia, “Overview of the protein- protein interaction annotation extraction task of Biocreative II”, Genome Biology, Vol. 9, 2008, p. 4.
- [73] P. Kordjamshidi, D. Roth, M. F. Moens, “Structured learning for spatial information extraction from biomedical text: bacteria biotopes”, BMC Bioinformatics, Vol. 16, 2015, pp. 129-129.
- [74] E. M. Voorhees, L. P. Buckland, “Proceedings of the Sixteenth Text Retrieval Conference”, NIST, 2007, pp. 500-274.
- [75] S. Riedel, D. McClosky, M. Surdeanu, A. McCallum, C. D. Manning, “Model combination for event extraction in BioNLP”, Proceedings of the BioNLP Shared Task 2011 Workshop, Association for Computational Linguistics, Portland, OR, USA, June 2011, pp. 51-55.
- [76] J. Bjerne, T. Salakoski, “TEES 2.1: automated annotation scheme learning in the BioNLP 2013 shared task”, Proceedings of the BioNLP Shared Task 2013 Workshop, Association for Computational Linguistics, Sofia, Bulgaria, August 2013, pp. 16-25.
- [77] Q. C. Bui, D. Campos, E. V. Mulligen, J. Kors, “A fast rule-based approach for biomedical event extraction”, Proceedings of the BioNLP Shared Task 2013 Workshop, Association for Computational Linguistics, Sofia, Bulgaria, August 2013, pp. 104-108.
- [78] Y. Chen, L. Xu, K. Liu, D. Zeng, J. Zhao, “Event extraction via dynamic multi-pooling convolutional neural networks”, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Vol. 1, 2015, pp. 167-176.
- [79] T. Wang, Q. Kema, “GPU Acceleration for Optical measurement”, SPIE Press, 2017.
- [80] J. Wang, H. Li, A.Y. Lin, H. Z. Yang, “Biomedical event trigger detection based on convolutional neural network”, International Journal of Data Mining and Bioinformatics, Vol. 15, No. 3, 2016, pp. 195-213.
- [81] Z. Han, J. Jiang, L. Qiao, Y. Dou, J. Xu, Z. Kan, “Accelerating Event Detection with DGCNN and FPGAs”, Electronics, Vol. 9, No. 10, 2020, p. 1666.

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