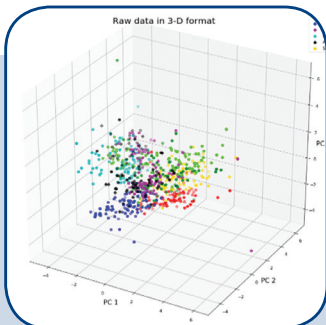
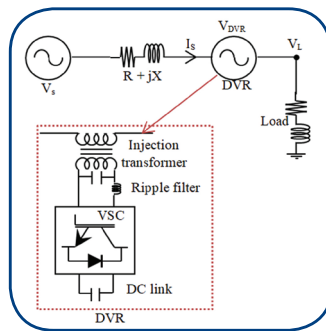
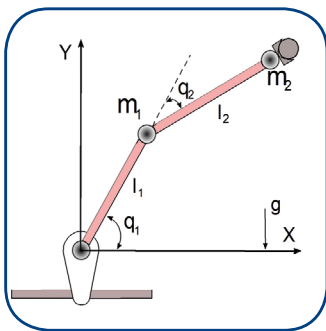


International Journal of Electrical and Computer Engineering Systems



```
contract SimpleContract {
  mapping( address => uint ) public balance;

  function donate( address to ) payable {
    balance[to] += msg.value;
  }

  function withdraw( uint amount ) public {
    if (balance[msg.sender] >= amount) {
      require(msg.sender.call.value(amount));
      balance[msg.sender] -= amount;
    }
  }

  function queryCredit( address to ) view returns (uint) {
    return balance[to];
  }
}
```



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 13, Number 9, 2022 | Pages 729 - 837

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
(IJECS)**

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

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 Secleanu
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

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

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

Journal is referred in:

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

Bibliographic Information

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

IJECS 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

Comparative Analysis between Elgamal and NTRU Algorithms and their implementation of Digital Signature for Electronic Certificate	729
<i>Original Scientific Paper</i>	
Bambang Harjito Tri Setyowati Ardhi Wijayanto	
A comparative study of hash algorithms with the prospect of developing a CAN bus authentication technique	741
<i>Original Scientific Paper</i>	
Asmae Zniti Nabih El Ouazzani	
Towards Auto Contract Generation and Ensemble-based Smart Contract Vulnerability Detection	747
<i>Original Scientific Paper</i>	
K. Lakshmi Narayana K. Sathiyamurthy	
An empirical study on English-Mizo Statistical Machine Translation with Bible Corpus	759
<i>Original Scientific Paper</i>	
Chanambam Sveta Devi Loitongbam Sanayai Meetei Bipul Syam Purkayastha	
Speaker Recognition Based on Mutated Monarch Butterfly Optimization Configured Artificial Neural Network	767
<i>Original Scientific Paper</i>	
Dhana Lakshmi Namburi Satya Sai Ram M	
Multimodal Behavioral Biometric Authentication in Smartphones for Covid-19 Pandemic	777
<i>Original Scientific Paper</i>	
Amitabh Thapliyal OP Verma Amioy Kumar	
Decision Support Machine - A Hybrid Model for Sentiment Analysis of News Headlines of Stock Market	791
<i>Original Scientific Paper</i>	
Kirti Sharma Rajni Bhalla	
Identifying and Classifying an Ovarian Cyst using SCBOD (Size and Count-Based Ovarian Detection) Algorithm in Ultrasound Image	799
<i>Original Scientific Paper</i>	
S. Jeevitha N. Priya	
ResViT: A Framework for Deepfake Videos Detection	807
<i>Original Scientific Paper</i>	
Wasim Ahmad Imad Ali Sahibzada Adil Shahzad Ammarah Hashmi Faisal Ghaffar	
Design of Super Twisting Integral Sliding Mode Control for Industrial Robot Manipulator	815
<i>Original Scientific Paper</i>	
Shankar J Gambhire D Ravi Kishore Malligunta Kiran Kumar Sushant N Pawar	
Design of High-Speed Dual Port 8T SRAM Cell with Simultaneous and Parallel READ-WRITE Feature	823
<i>Original Scientific Paper</i>	
Shourin Rahman Aura S. M. Ishraqul Huq Satyendra N. Biswas	
Comparative Performance of DVR and STATCOM for Voltage Regulation in Radial Microgrid with High Penetration of RES	831
<i>Case Study</i>	
Ritika Gour Vishal Verma	
About this Journal	
IJECES Copyright Transfer Form	

Comparative Analysis between Elgamal and NTRU Algorithms and their implementation of Digital Signature for Electronic Certificate

Original Scientific Paper

Bambang Harjito

Sebelas Maret University,
Faculty of Math and Natural Science, Department of Informatics
Surakarta, Indonesia
bambang_harjito@staff.uns.ac.id

Tri Setyawati

Sebelas Maret University,
Faculty of Math and Natural Science, Department of Informatics
Surakarta, Indonesia
trisetyawati11@student.uns.ac.id

Ardhi Wijayanto

Sebelas Maret University,
Faculty of Math and Natural Science, Department of Informatics
Surakarta, Indonesia
ardhi.wijayanto@staff.uns.ac.id

Abstract – The emergence of electronic certificates, which are official documents in the form of digital files transmitted via the internet, facilitates the exchange of information. However, internet use has risks, such as data theft for fabricating and modifying information. This problem can be solved by applying a digital signature. The main concern in this research is how to perform a comparative analysis between asymmetric cryptographic Elgamal and NTRU (Nth-Degree Truncated Polynomial Ring) algorithms and their implementation of a digital signature as an effort to improve information security in electronic certificates. The stages of the research method are divided into the key generation process, signing, and verification. In the signing and verification process, the SHA-512 hash function is also used for hashing messages to be encrypted-decrypted and QR Code as the signature. Comparison of performance of NTRU with Elgamal algorithms required running at a pdf extension with security levels 80,128,192, 256 bits will be obtained from the templates.office.com website. The results obtained that the El Gamal algorithm is better than the NTRU algorithm, but at a higher security level, the NTRU algorithm is better than the Elgamal algorithm. In the verification experiment that has been carried out, it can be concluded that by using SHA-512 as a hash function, the N parameter used for NTRU must be greater than or equal to 512 to avoid error results from verification.

Keywords: NTRU, Elgamal, Electronic Certificate, Digital Signature, SHA-512, QR Code

1. INTRODUCTION

Electronic certificates are official documents in the form of digital files transmitted via the internet, where the internet itself is vulnerable to theft and falsification of information, such as the fabrication or modification of information [1,2,3]. To increase the security of electronic certificates. A security system in the form of digital signatures is applied to the electronic certificates. The digital signature is a cryptographic value that depends on the message's content and the message's

sender. So that different messages with the same sender will have different digital signatures [4,5,6,7].

This study aims to implement a digital signature on an electronic certificate using the NTRU and Elgamal algorithm at security levels 80,128,192, and 256 bits to see whether NTRU is better or Elgamal is better. NTRU algorithm is an asymmetric algorithm. Asymmetric algorithms have different keys during the encryption and decryption process, namely the public and private keys. The public key is the key that is published and may be known by ev-

everyone, while the private key is a key that is kept secret and may only be understood by one person [8,9]. The NTRU algorithm's security level lies in the use of polynomials during the operation process, as well as the difficulty of finding a short vector of a lattice [10,11]. The level of security of the Elgamal algorithm lies in the difficulty of calculating discrete logarithms [12]. For comparison, the Elgamal algorithm is used at security levels 80,128,192, and 256 bits to see which algorithm has better performance. Elgamal's algorithm selection is the comparison because the study [37] shows that Elgamal is a better probabilistic algorithm than RSA and has a difficulty level that lies in discrete logarithm calculations and its ability to solve fundamental distribution problems. Besides that, there are still very few studies comparing the NTRU algorithm with the algorithm Elgamal.

This research aims to implement a digital signature schemes using the NTRU and Elgamal algorithm for electronic certificates, then analyze the running time on the NTRU and Elgamal algorithms based on the process generate keys, signing, and verifying, as well as analyze the results of electronic certificate verification.

2. RELATED WORK

The NTRU algorithm is a fast and lightweight public key algorithm to provide end-to-end security that can be used to improve document security standards with better encryption and decryption than the RSA and ECC algorithms [13,14,15]. In another study, to increase document security and facilitate the validation process, the use of digital signatures using the RSA algorithm [16,17] and SHA-512 algorithms can be applied where the method used is to generate a public key and a private key with RSA. The signing process is carried out by encrypting the message digest generated from the message hashing process with SHA-512 and then verifying electronic documents by matching the results of document decryption and SHA-512 hashing of documents [18,19,20,21] in a similar study [3]. With SHA-3 hashing function and super encryption combination of RSA and AES, with QR-Code scheme to accommodate the signature code. In this implementation, the certificate will be signed with the SHA-3 hashing process sequence, encrypted with RSA, encrypted with AES, and ends by embedding the QR-Code that has been generated from the AES encryption results on the electronic certificate.

3. THEORY USED

This section discusses the theoretical background to analyze the comparison between the NTRU and Elgamal Algorithms and the implementation of digital signatures on electronic certificates.

3.1. DIGITAL SIGNATURE

Digital Signature is a means used to view authentication on digital messages, both messages transmitted through communication channels and electronic

documents; what is meant by digital signatures are signatures that have a cryptographic value that depends on the content of the message and the sender of the message, so that the message different ones with the same sender will have different digital signatures [14]. The term direct digital signature refers to a digital signature scheme that only involves the communicating party (sender, receiver). Digital signature schemes are similar to asymmetric cryptographic systems in that they involve public and private keys and run an algorithm that uses these keys to sign and verify [22,23].

3.2. NTRU (NTH-DEGREE TRUNCATED POLYNOMIAL RING UNITS) ALGORITHM

NTRU uses addition and multiplication operations in line with Ring, which is an algebraic object that has two operations, addition, and multiplication, which are related via the distributive law [22, 23], NTRU works with rings. An element will be written as a polynomial or vector according to Equation (1).

$$F = \sum_{i=0}^{N-1} F_i x^i = [F_0, F_1, \dots, F_{N-1}] \quad (1)$$

Key Generation: Choose two polynomials $f \in L_f$ $f \in$ and $g \in L_g$. Polynomial f must meet the additional requirement that it has inverse modulo q and inverse modulo p . This inverse can be expressed F_p and F_q so that the result is:

$$F_q * f \equiv 1 \pmod{q} \quad (2)$$

$$F_p * f \equiv 1 \pmod{p} \quad (3)$$

Next calculate h with Equation (4)

$$h \equiv pF_q * g \pmod{q} \quad (4)$$

Where h is a polynomial that functions as a public key and a polynomial f, fp as private keys.

NTRU encryption: Selects m as messages from a set of plaintexts L_m . Then randomly choose the polynomial L_ϕ and use the public key h to compute the encrypted message e by Equation (5)

$$e \equiv \phi * h + m \pmod{q} \quad (5)$$

Polynomial e is an encrypted message that will be sent to the recipient of the message.

NTRU decryption: In the decryption of a received e-message, the process is carried out using the private key f to calculate the value of a with Equation (6).

$$a \equiv f * e \pmod{q} \quad (6)$$

Where the coefficient a is in the interval from $q/2$ to $-q/2$. Now with a as a polynomial with integer coefficients and a private key, F_p , which can be used to recover m messages with Equation (7)

$$m \equiv F_p * a \pmod{p} \quad (7)$$

3.3. ELGAMAL ALGORITHM

Key Generation: Elgamal has a parameter of key size, which will later be used to determine positive prime numbers and integers that are primitive roots of p . To generate a public key and a private key is done by choosing a random number x , provided that than calculate the value of y with Equation (8)

$$y = g^x \pmod{p} \quad (8)$$

The result of key generation is in the form of private key x and public key y, g , and p .

Elgamal Encryption: Before performing the encryption process, first declare the message as an integer m and must lie in the range $[0, p-1]$. For large m , divide m , into smaller blocks so that each block represents a value in the range $[0, p-1]$. The encryption steps are as follows:

1. Choose a random number k , provided that $1 \leq k \leq (p - 1)$
2. Encrypt message m into value pairs (a, b) with the Equation:

$$a = g^k \pmod{p} \quad (9)$$

$$b = y^k m \pmod{p} \quad (10)$$

The pairs a and b are the ciphertext for message m . So, the ciphertext size is twice the size of the plaintext.

Elgamal's description: For decryption, the private key x is used to decrypt a and b into plaintext m with Equation [11].

$$m = b(a^x)^{-1} \pmod{p} \quad (11)$$

3.4. SHA-512 (SECURE HASHING ALGORITHM)

SHA (Secure Hashing Algorithm) is designed by the National Security Agency (NSA). SHA security is based on the fact that a birthday attack on a digest of n bits results in a collision with a work factor of about $2n/2$ [24]. SHA-512 is one of the results of the revision of the FIPS standard in 2002, which defines three new versions of SHA, with hash values of 256, 384, and 512-bits long, known as SHA-256, SHA-384, and SHA-512. Collectively, these hash algorithms are known as SHA-2. This new version has the same basic structure and uses the same types of binary logical operations and modular arithmetic as SHA-1 [12].

3.5. QR CODE

QR Code is a two-dimensional matrix symbology with a position detection pattern at three angles initially designed for very high-speed reading and Omnidirectional reading. The QR Code was developed to increase the speed of reading complex structured 2D barcodes. Other QR Code features are bulk data capacity, high data density, and selectable levels of error cor-

rection capability [25, 26,27,28]. QR codes store data using a graphical representation. The essence of this representation is based on the arrangement of several simple geometric shapes on a fixed space [29,30,31,32].

4. PROPOSED WORK

This section provided an overview of our solution comparative analysis of the Elgamal and NTRU Algorithm, and the Implementation of digital signatures on an electronic certificate.

Fig.1 shows the comparative analysis model of the digital signature in the NTRU and Elgamal algorithms in carrying out the signing and verifying process. The comparative analysis model consists of two processes: (1) the signing process using both NTRU and Elgamal Algorithms and (2) the verifying process using both NTRU and Elgamal Algorithm. The process begins with the user who selects the electronic certificate with a pdf extension file. This file can be called plaintext.

- a) The signing process using both NTRU and Elgamal Algorithms

The signing process begins with (1) calculating the hash value of the certificate file with the SHA-512 hash function, which produces a message digest (m). (2) these results are then encrypted with the Private Key and produce a message digest cipher (3) The cipher message digest is then stored in the form of a file pythons. (4) Generate a QR Code with the cipher message digest file address as the data. QR Code generation is done using a QR Code Generator. (5) Embed the QR Code as a signature into the certificate file. (6) Sending certificate file + QR Code to recipient.

- b) The In the verifying process using both NTRU and Elgamal Algorithm

The verifying process is executed with (1) Look for the signature in the form of a QR Code, which is contained in the certificate file, then separate the QR Code from the certificate. (2) Calculate the hash value of the certificate file, and generate a message digest (m'). (3) Decode the QR Code to get the data in it, which is the address of the cipher message digests file. (4) Based on the address obtained, then look for the cipher message digest file to get the cipher message digest. (5) Decrypt the cipher message digest with the Public Key, which results in a message digest (m). (6) Comparing m and m' . If $m = m'$, then it can be concluded that the certificate file is "Valid". Meanwhile, if it is not the same, it can be concluded that the certificate file is "Invalid"

Testing is done by running the system to see if the system has running according to its function or not, by doing signing and verifying experiments document. The certificate document will be subject to a signing process and generate a new certificate document complete with a digital signature in it. This document will then be subject to two treatments, (1) The certificate document does not subject to any content changes, so

when the verification process is carried out with the system, it will display a result indicating that the certificate is correct or valid. (2) The certificate document is subject to content changes so that when the verification process is carried out with the system, it will display the

following results indicating that the certificate is incorrect or invalid. After testing, system test results data in the form of running time during the key generation, sign, and verify processes. The data is then analyzed to see whether NTRU is better or ElGamal is better.

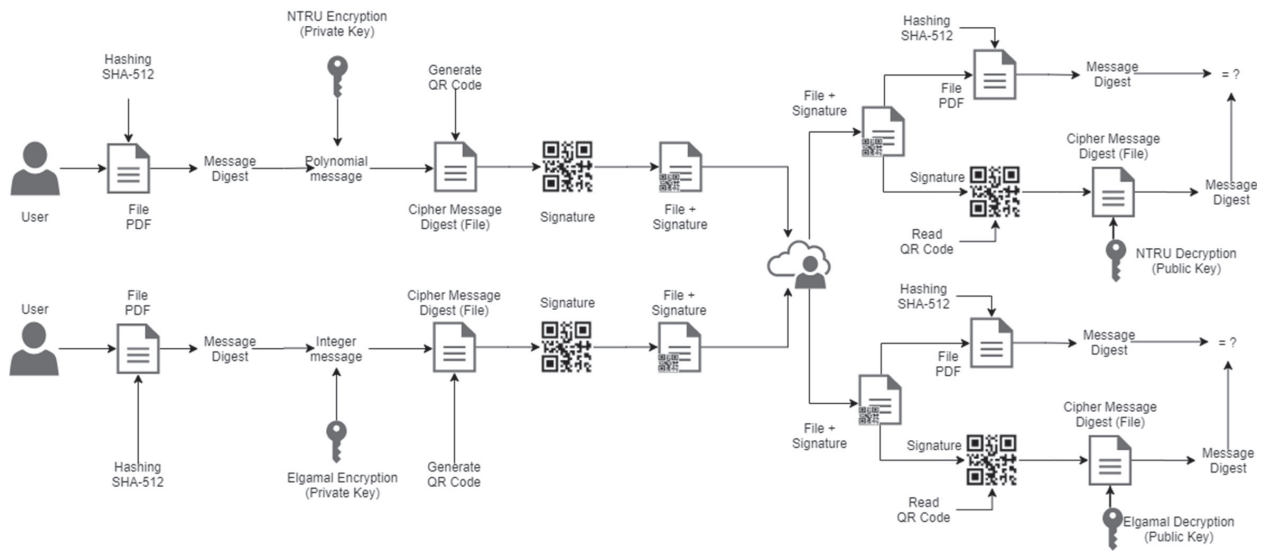


Fig. 1. Comparative Analysis of ElGamal and NTRU Algorithm and Implementation in the electronic certificate

5. RESULT AND DISCUSSION

In this section, perform a comparative analysis of the NTRU and ElGamal Algorithm and an Implementation in the electronic certificate.

5.1. DATA COLLECTION

The data for this research is a certificate with a pdf extension obtained from the templates.office.com website. The certificate can be depicted in Fig 2.

The performance of NTRU with ElGamal required data showing that the NTRU with such parameters will be comparable to ElGamal with so many bits. The data which can be used for this purpose is the security level of the NTRU algorithm with the ElGamal algorithm [33, 34, 35, 36]. The data can be seen in Table 1.

Table 1. Security Level Elgamal and NTRU

Security Level (bits)	NTRU	Elgamal (bits)
80	251	1024
128	397	3072
192	587	7680
256	787	15360

5.2. NTRU IMPLEMENTATION OF THE DIGITAL SIGNATURE SCHEME

NTRU Key Generation: The main parameters of the NTRU algorithm are integers N , p , and q , and the four sets L_f, L_g, L_m, L_ϕ polynomial of degree $N-1$ with integer coefficient. The integers p and q do not have to be prime, but provided that $\gcd(p, q) = 1$, and q will always be much greater than p . NTRU works with rings polynomial $R = \mathbb{Z}[X]/(X^N - 1)$.

For example, the low parameter is chosen to facilitate the ease of its calculation. The selected parameters are $N = 251, p = 3$, and $q = 2048$. Then randomly determine the polynomials f and g as follows:

$$g = x^{*247} + x^{*245} - x^{*244} + x^{*240} + x^{*235} + x^{*225} + x^{*221} + x^{*220} + x^{*219} - x^{*217} + \dots - x^{*36} - x^{*34} + x^{*31} - x^{*30} + x^{*29} + x^{*27} + x^{*21} + x^{*19} - x^{*16} + x^{*15} - x^{*13} - x^{*7} - x^{*2} - x - 1$$

$$f = x^{*250} + x^{*249} - x^{*248} - x^{*246} + x^{*244} - x^{*243} - x^{*242} + x^{*241} - x^{*239} + x^{*238} - x^{*237} - \dots - x^{*17} + x^{*16} - x^{*15} + x^{*14} - x^{*13} - x^{*12} - x^{*9} - x^{*8} + x^{*6} + x^{*5} + x^{*4} - x^{*3} + x^{*2} - 1$$

Then calculate the inverse of $f \text{ mod } p$ and $f \text{ mod } q$. The results obtained are

$$f_p^{-1} = x^{*250} + x^{*249} + x^{*247} + x^{*245} + x^{*244} + x^{*243} - x^{*241} - x^{*240} - x^{*239} + x^{*236} - x^{*235} + x^{*234} - \dots + x^{*22} + x^{*21} + x^{*19} - x^{*18} + x^{*17} + x^{*15} - x^{*14} + x^{*10} + x^{*8} + x^{*7} + x^{*6} - x^{*5} + x^{*4} - x^{*3} - x$$

$$f_q^{-1} = -919x^{*250} - 141x^{*249} + 376x^{*248} + 556x^{*247} + 275x^{*246} + 150x^{*245} + 201x^{*244} + \dots + 249x^{*9} - 199x^{*8} + 805x^{*7} + 384x^{*6} + 216x^{*5} + 864x^{*4} + 819x^{*3} - 696x^{*2} + 686x + 1019$$



Fig. 2. Certificate Data with a pdf extension

The last step is the public key calculation

$$h = 1013x^{250} + 684x^{249} + 737x^{248} + 355x^{247} - 113x^{246} - 991x^{245} - 652x^{244} + 473x^{243} - \dots - 571x^8 - 531x^7 - 414x^6 - 466x^5 + 847x^4 - 214x^3 - 926x^2 - 1014x + 718$$

The result of generating the NTRU key is stored in a python file, as shown in Fig. 3, and is used in the signing and verifying process as input.



Fig. 3. NTRU Key File

NTRU Signing: The signing process requires the input of a certificate file with extension pdf and a key file with extension npz. The message in the certificate file will be hashed with the SHA-512 hash function and produce a message digest. These results are then encrypted using the NTRU algorithm with completion $e \equiv r * h + m; (mod q)$ so that the results are in the form of ciphertext. The resulting ciphertext will be stored in a file and the address of the file is then used as data in QR Code generation, by utilizing the QR Code module from Python and acts as a signature.

To be clear, assume that the message is already in the form of polynomial.

$$m = x^{509} + x^{507} + x^{506} + x^{505} + x^{504} + x^{503} + x^{502} + x^{499} + x^{498} + x^{493} + x^{490} + x^{489} + \dots + x^{38} + x^{36} + x^{33} + x^{31} + x^{28} + x^{26} + x^{14} + x^{12} + x^{11} + x^{07} + x^{06} + x^{05} + 2x + 2$$

Then calculate a random polynomial r of degree 251. Assume that the result is:

$$r = x^{202} + x^{186} + x^{183} + x^{180} + x^{176} + x^{169} + x^{167} + x^{163} + x^{151} + x^{146} + x^{135} + x^{127} + x^{126} + x^{123} + x^{121} + x^{114} + x^{112} + x^{109} + x^{107} + x^{100} + x^{97} + x^{94} + x^{93} + x^{92} + x^{88} + x^{84} + x^{83} + x^{82} + x^{80} + x^{75} + x^{67} + x^{66} + x^{64} + x^{60} + x^{59} + x^{58} + x^{50} + x^{46} + x^{45} + x^{38} + x^{37} + x^{36} + x^{34} + x^{33} + x^{31} + x^{28} + x^{26} + x^{14} + x^{12} + x^{11} + x^{07} + x^{06} + x^{05}$$

Based on the encryption formula, the encrypted message value is obtained with the public key in the previous example.

$$e = -39x^{250} + 981x^{249} + 124x^{248} - 90x^{247} + 238x^{246} - 129x^{245} - 147x^{244} - 217x^{243} - \dots - 173x^{09} + 808x^{08} - 981x^{07} - 4x^{06} + 952x^{05} + 962x^{04} + 472x^{03} + 935x^{02} + 185x - 946$$

The final result of the signature process (with parameters $N = 587, p = 3,$ and $q = 2048$) is a pdf certificate file, as shown in Fig 4.



Fig. 4. Signed Certificate File

During the signing process, the data, which was originally a plaintext message, was processed to become a signature. An example of the process of changing the data can be seen in Table 2.

Table 2. Data Changes During NTRU Signing

Process	Results
Read the contents of the certificate file	CERTIFICATE OF TRAINING This certifies that Tengiz Kharatishvili has successfully completed training in video publishing Rowan Murphy, Sr. Videographer June 04, 20XX Date
Hashing messages with SHA – 512	b0f741a08c914065d146114e6f946b 50449ca05a87753cf014cb572eec68 dbb5172457fa9e49a9f188a1a377ed 9047eb1be64d1d61c27f4327fc56f8 9d85fba4
Converting message to polynomial form	Poly($x^{511} + x^{510} + x^{509} + x^{508} + \dots + x^{22} + 2x + 2, x, \text{domain}='ZZ'$)
Encryption result	Poly($38x^{586} - 28x^{585} - \dots - 26x^{33} - 38x^{32} - 29x + 9, x, \text{domain}='ZZ'$)
Save the encryption result into file	encrypted_220512_191310.npz

Generate QR-Code



Verifying NTRU: The verification process is initiated by inputting the signed certificate file and key file. The certificate file will then be carried out by two different processes, namely (a) the process of hashing the contents of the certificate file, which produces a message digest, and (b) the process of decrypting the ciphertext using the completion of the NTRU algorithm $d \equiv f_p^{-1} * [f * e]_q (mod p)$. The decryption process can be done by first decoding the QR Code to obtain the ciphertext file address.

More specifically, take the values of $e, f,$ and $fp-1$ using the previous encryption calculations. Then use the private key f to calculate the value of d with the formula until you get the result:

$$d = x^{250} + x^{248} + x^{247} + x^{244} + x^{242} + x^{241} + x^{240} + x^{239} + x^{238} + x^{236} + x^{234} + x^{233} + \dots + x^{22} + x^{21} + x^{17} + x^{14} + x^{13} + x^{12} + x^{11} + x^{09} + x^{08} + x^{06} + x^{05} + x^{04} + x^{03} + x^{02}$$

An example of the process of changing the data can be seen in Table 3.

Table 3. Data Changes During NTRU Verifying

Process	File	Signature
Processed data	CERTIFICATE OF TRAINING This certifies that Tengiz Kharatishvili has successfully completed training in video publishing Rowan Murphy, Sr. Videographer June 04, 20XX Date	 QR Code Decode Result: encrypted_220512_191310.npz

	Message Digest:	Decryption Result:
	b0f741a08c914065d	b0f741a08c914065
	146114e6f946b5044	d146114e6f946b50
	9ca05a87753cf014cb	449ca05a87753cf01
Results	572eec68dbb517245	4cb572eec68dbb51
	7fa9e49a9f188a1a37	72457fa9e49a9f188
	7ed9047eb1be64d1	a1a377ed9047eb1b
	d61c27f4327cf56f89	e64d1d61c27f4327f
	d85fba4	c56f89d85fba4

Verification result	Valid Certificate
---------------------	-------------------

5.3. ELGAMAL IMPLEMENTATION OF THE DIGITAL SIGNATURE SCHEME

Elgamal Key Generation: Elgamal's algorithm uses a parameter in the form of a key size, which will later be used to determine the positive prime number p and the integer q , which is the primitive root of p . More specifically, for example, a low parameter will be chosen so that calculations can be carried out easily. The selected parameter is a key length of 1024 bits.

Determine the positive prime number p and the primitive root integer (of p) g , as follows:

p :

```
12489857221665811115722087408354707200904245054949915548595594703477566992936482827403733
844778047412215234604025630607003873323702143405841378919941167032148636454799401302956338
429661580291454927374628123190632069773463945113846463535827792773210978304435023477920819
1301968003951971300258281822235019146787
```

g :

```
634187109953871722414031920698055381971645540161670912639191994243896878671751031164050423
516381048310300668121586108891709476922650706546659951979204975000220138786820104549193674
989313904985155285516103341031838636861024714927332979007191006009952392275449498672360104
06029917652700377892076852541591702248
```

After that, a random number x that meets the conditions can be determined:

x :

```
91491006234848614245179977624682077925335183449227099755560879669520524960131412926704708
34735786780210541158300926247902230874565092117625576634432767085655123007686017958418385
032352969005034382561312282392552388261175928361869315037599080053753237342233076253295610
3199301369876073313607616134516448307
```

Then calculate the value of the public key y using the formula until it gets the result:

y :

```
6513528886376233423156589926619645695596911824928693298285511568454874973191493243304338
387323226204977601743100432410851026706922059501650769220930864413183397072789375870617976
6387255086140429860544959324431905194105832001038932251989629771258967253252343837800990
05128460592394696256845127484968882332
```

The result of the Elgamal key generation is stored in a python file, as shown in Fig. 5, and is used in the signing and verifying process as input.



Fig. 5. Key File

Elgamal Signing: The signing process requires the input of a certificate file with a pdf extension and a key file with npz extension. Messages in the certificate file are hashed with SHA-512 and generate a message digest. These results are then encrypted using the Elgamal algorithm with completion $a = g^k \pmod{p}$ and $b = y^k m \pmod{p}$.

So that the results are in the form of ciphertext. The resulting ciphertext will be stored in a file and the address of the file is then used as data in QR Code generation by utilizing the QRcode module from Python and acts as a signature.

More specifically, for example, assume that the message you want to send is already in the form of an integer:

m :

```
[14538442250038144105231320892152646169614843557636457374901122892002070321499286938540856
146613390051004182394475048657577252138832684349648204124476643815169212696928541858932923
96867705009816910287787229500894085499216436358018904406817756050209176272038926760434163
5003563025971688376388564955828518655,
277051609433009417896108549721465843810604790226550854273301672281256388876166909189707041
387833196283598744810191333262733937741703000683156540838356623441981685851286315026748872
72242620366028390917958819356794477409021229284543785907985788707577484326733185017925868
63442741449112785962979225229952101, 51]
```

Then determine the random number k :

k :

```
413595623204417673776342832888503224125472768031594611495484770241284420055844750071727967
000160416200371528602432811785808733490802274781083247710926862789684457461269585324098123
751015437125155307002915256627810925348141437771779326514895333116231344373179558851453931
72700788991929703414529595208367799321
```

By using the values of g , p , and y obtained in the previous generate key, then calculating the values of a and b with the formula until the following results are obtained.

a :

```
704798831812903730767592840850362047259304136001732045823020489237667079342633510409537862
74254348456281394440796452980960199763720656329447552526439670077386342709719064201262707
090770638523139374257056105282585107662663863042102239634363113079173711537825725359070539
66206171964146582994168138422591559688
```

b :

```
381469081370942410420486521225778003348021939250172661701837602867678272072201630372568428
102398400480159146431599276831352445023784702785633990400058547222602018490187243485771195
827375432340619371181532871649931584390177232329400813336145567897002704448237950502335543
92401223836825760672707159682353497160
```

The final result of the signature process is a pdf certificate file, as shown in Fig. 6.



Fig. 6. Signed (Elgamal) Certificate File

During the signing process, the data, which was initially a plaintext message, was processed to become a signature. An example of the process of changing the data can be seen in Table 4.

Table 4. Data Changes During Elgamal Signing

Process	Results
Read the contents of the certificate file	CERTIFICATE OF TRAINING This certifies that Tengiz Kharatishvili has successfully completed training in video publishing Rowan Murphy, Sr. Videographer June 04, 20XX Date
Hashing messages with SHA – 512	b0f741a08c914065d146114e6f946b50449ca05a87753cf014cb572ecc68dbb5172457fa9e49a9f188a1a377ed9047eb1be64d1d61c27f4327fc56f89d85fba4
Convert message digest to integer	2688241...950207, 2660811...905845, 52
Determine the random number k	6791467875486908727...72474959180919163072209829196994526
Encryption result	82875883331547...83931630431734, 56002328522083...40235165644727, 79852418918538...58497092617061, 31666415267698...84650783518716, 11618727619395...39185616194775, 37640656195438...72911613317981
Save the encryption result into file	keyElgamal_220512_203003.npz
Generate QR-Code	

Verifying Elgamal: The verification process begins with inputting a signed certificate file and a key file. The certificate file will then be carried out by two different processes, namely (a) the process of hashing the contents of the certificate file, which produces a message digest, and (b) the process of decrypting the ciphertext using the completion of the NTRU algorithm $m = b(a^x)^{-1} \pmod p$. The decryption process can be done by first decoding the QR Code to obtain the ciphertext file address.

More specifically, use the private keys x , ciphertext $[a, b]$, and p from the previous calculation, then calculate the value of m with the formula until the results are obtained

m:
145384422500381441052313208921526461696149435576364573749011228920020703214992869385408561
466133900510041823944750486575772521388326843496482041244766438151692126969285418589329239
68677050098169102877872295008940854992164363580189044068177560502091762720389267604341635
803563025971688376388564955828518655,
277051609433009417896108549721465843810604790226550854273301672281256388876166909189707041
387833196283598744810191333262733937741703000683156540838356623441981685851286315026748872
72242620366028390917958819356794477409021229284543785907985788077577484326733185079125868
634427414491127859629792255229952101

An example of the process of changing the data can be seen in Table 5.

Table 5. Data Changes During Elgamal Verifying

Process	File	Signature
Processed data	CERTIFICATE OF TRAINING This certifies that Tengiz Kharatishvili has successfully completed training in video publishing	

Processed data	Rowan Murphy, Sr. Videographer June 04, 20XX Date	QR Code Decode Result: keyElgamal_220512_203003.npz
Results	Message Digest: b0f741a08c914065d146114e6f946b50449ca05a87753cf014cb572ecc68dbb5172457fa9e49a9f188a1a377ed9047eb1be64d1d61c27f4327fc56f89d85fba4	Decryption Result: b0f741a08c914065d146114e6f946b50449ca05a87753cf014cb572ecc68dbb5172457fa9e49a9f188a1a377ed9047eb1be64d1d61c27f4327fc56f89d85fba4
Verification result	Valid Certificate	

5.4. COMPARISON OF NTRU AND ELGAMAL ALGORITHM

Key Generation: Longer keys will provide higher security but consume more computational time, so the value of safety and speed will be inversely related. Table 6 shows the running time results for key generation on the Elgamal and NTRU algorithms.

Table 6. Running Time of NTRU and Elgamal Key Generation

Security Level	Algorithm	Trial (seconds)					Average
		1	2	3	4	5	
Low	NTRU	97.977	95.85	95.599	95.867	114.09	99.8766
	Elgamal	5.913	1.324	2.698	25.212	119.116	30.8526
Standard	NTRU	274.4	282.074	275.681	250.869	252.186	267.042
	Elgamal	133.610	4143.273	384.657	1712.502	282.778	1331.364
High	NTRU	532.952	527.532	539.384	684.279	532.874	563.4042
	Elgamal	3875.925	6818.925	7651.898	2647.856	2641.153	4727.151
Highest	NTRU	1255.205	1154.33	1106.095	1024.373	1026.911	1113.3828
	Elgamal	>9 Hours (32400)	>9 Hours (32400)	>9 Hours (32400)	>9 Hours (32400)	>9 Hours (32400)	> 9 Hours (32400)

From the Table 6, it can be described in graphical form, as shown in Fig. 7.

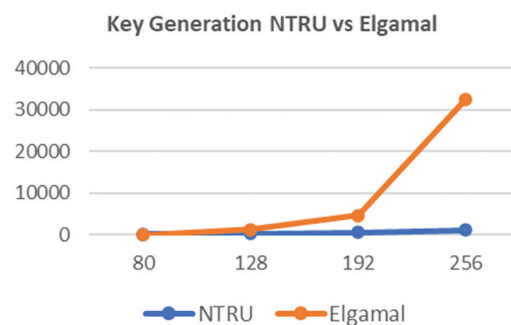


Fig. 7. Running Time of NTRU and Elgamal Key Generation

From Fig. 7, the computation time required to generate an NTRU key at a low-security level is 3x slower than Elgamal, but at a standard and high NTRU security level it is almost 5x and 8x faster than Elgamal

Signs: The time required to sign the file using the two algorithms is compared to evaluate the performance of the proposed system. The running time of the signing process with the NTRU and Elgamal Algorithms in five trials, results are shown in Table 7.

Table 7. Running Time Signing NTRU and Elgamal

Security Level	Algoritma	Trial (seconds)					Average
		1	2	3	4	5	
Low	NTRU	6.969	6.242	6.138	5.967	6.242	6.3116
	Elgamal	0.62	0.606	0.649	0.565	0.527	0.5934
Standard	NTRU	11.975	12.004	11.829	10.601	10.353	11.3524
	Elgamal	2.505	2.281	2.226	2.218	2.123	2.2706
High	NTRU	20.529	20.401	20.5	24.45	20.725	21.321
	Elgamal	28.23	27.072	31.958	34.558	31.604	30.6844
Highest	NTRU	50.471	45.614	46.38	40.799	40.158	44.6844
	Elgamal	-	-	-	-	-	-

Table 7 shows graphs of running time for the signing process using the Elgamal and NTRU algorithms, as shown in Fig. 8. Experimental running time on Elgamal with the highest security cannot be carried out due to a failure in the key generation process which cannot generate a public key and a private key, so the process cannot be continued.

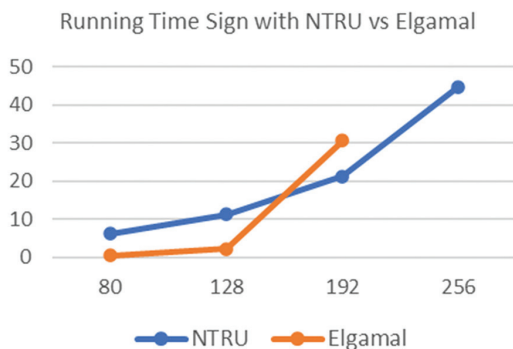


Fig. 8. Running Time Signing NTRU and Elgamal

From Fig. 8, the time for signing using Elgamal is faster than NTRU when the experiment is carried out at a low-security level and standard security. At the same time, at a higher security level, namely high safety, and highest security, the signing process using the NTRU algorithm requires faster time if compared to Elgamal's algorithm. So, at a higher level of protection, it can be said that the signing process using the NTRU algorithm is faster and safer than using the Elgamal algorithm. From Fig 8, the computation time required to sign NTRU at a high-security level is almost 1.5x faster than Elgamal.

Verify: The NTRU cryptosystem significantly produces faster average speeds than Elgamal when the key size is increased. The time required to sign the file using the two algorithms is compared to evaluate the performance of the proposed system. Table 8 shows the results of running time verifying NTRU and Elgamal

Table 8. Running Time Verifying NTRU and Elgamal

Security Level	Algoritma	Trial (seconds)					Average
		1	2	3	4	5	
Low	NTRU	8.578	7.815	7.538	7.619	8.121	7.9342
	Elgamal	0.527	0.648	0.555	0.54	0.531	0.5602
Standard	NTRU	20.843	20.985	22.962	19.06	18.214	20.4128
	Elgamal	3.614	3.610	3.458	3.443	3.509	3.5268
High	NTRU	39.782	39.462	40.5	50.229	40.109	42.0164
	Elgamal	64.338	70.588	59.873	61.728	61.132	63.5318
Highest	NTRU	101.122	83.753	84.22	77.998	79.882	85.395
	Elgamal	-	-	-	-	-	-

Table 8 shows graphs of running time Verifying NTRU and Elgamal Cryptosystem, as shown in Fig. 9.

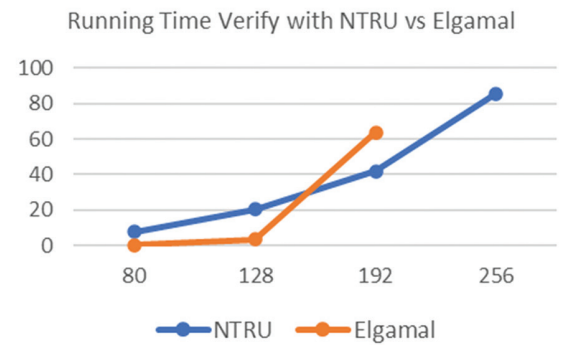


Fig. 9. Running Time Verifying NTRU and Elgamal

From Fig. 9, the verification process with the Elgamal algorithm for low-security levels, Elgamal is superior to NTRU. However, the security level is at a higher security level the NTRU is found to be faster than the Elgamal algorithm.

5.5. DIGITAL SCHEMATIC TESTING

NTRU: The test is carried out by running a digital signature scheme using the NTRU algorithm, against the same certificate file, with the final result as an "Invalid" or "Valid" certificate statement. It can be shown in Table 9.

Elgamal: The test is carried out by running a digital signature scheme using the Elgamal algorithm, against the same certificate file, with the final result in the form of an "Invalid" or "Valid" certificate statement. . It can be shown in Table 9.

Table 9 shows that in the Elgamal scheme, the verification results on certificates without changes at all security levels have "Valid" verification results, and certificates with changes have "Invalid" results. In the NTRU scheme, the verification results on the certificate with changes in all N-parameter tests result "Invalid", this happens because a difference in the contents of the certificate will result from a new message digest m (from the hashing process) which during the comparison process the value of m and m' (message digest from the decryption process) will be different and cause the verification results to be invalid.

Table 9. Testing the Digital Signature Scheme with NTRU

Security level (bits)	NTRU			Elgamal		
	N	Certificate Verification Results (Without Changes)	Certificate Verification Results (With Changes)	bits	Certificate Verification Results (Without Changes)	Certificate Verification Results (With Changes)
80	251	Invalid	Invalid	1024	Valid	Invalid
128	397	Invalid	Invalid	3071	Valid	Invalid
192	587	Valid	Invalid	7680	Valid	Invalid
256	787	Valid	Invalid	15360	-	-

However, Table 9 shows that the results of certificate verification without changes with parameter values NTRU N – 587 and N-787 show “Valid” results. The verification results do not change because the length of the polynomial ring R can include the size of the original message polynomial with the highest degree of 511. When decrypted, the decrypted polynomial will not be truncated because of the polynomial ring rule R . So that from the beginning to the end of the decryption process, the polynomial length of the processed message will not be truncated and intact. Meanwhile, the N – 251 and N -397 tests show “Invalid” results, which can occur due to the use of the SHA-512 hash function and the N parameter value that affects the length of the polynomial ring $R = \mathbb{Z}[X]/(X^N - 1)$. When the original message is hashed with SHA-512, which is then converted into polynomial form, the resulting polynomial will have a maximum length of 511 degrees. Meanwhile, during the encryption and decryption process, the size of the polynomial will follow the rules $R = \mathbb{Z}[X]/(X^N - 1)$. If the N parameter value used is 251, then the highest degree of the applicable polynomial is N-1 or 250. When the original message polynomial has the highest degree of 511 while the message decryption polynomial only has the highest degree of 250 because it follows the ring polynomial R . Then the message digest generated from the decryption process is different from the message digest from hashing the original message. So that the verification results show the same "Invalid" results, namely "Invalid" for the same reason.

6. CONCLUSION

From the problems encountered, the proposed problem-solving solutions, as well as the experiments carried out. It can be concluded that with the application of the NTRU and Elgamal Algorithms in the digital signature scheme, based on the comparison of running time in the key generation, sign, and verify processes, it shows that when the security level is low, NTRU is slower than Elgamal, but at the high-security level, NTRU is faster than Elgamal, which is 1.4x faster in the signing process and 1.5x faster in the verification process. So it can be said that NTRU, at a higher level of security, has faster when compared to Elgamal. In addition, the results of the certificate verification test with NTRU and Elgamal on the digital signature have been tested to be safe. This is proven by the testing process where data

modification is carried out in the certificate document, and the program manages to find out and shows the results "Invalid ". In the certificate document without modification, the program shows the result "Valid", but this result does not apply to NTRU N-251 and NTRU N-397 and shows the result "Invalid," which should be "Valid" this can happen because of the role of bit length The SHA used is SHA-512.

7. REFERENCES:

- [1] N. Yanti et al. Implementation of Advanced Encryption Standard (AES) and QR code algorithm on digital legalization system. in The 3rd International Conference on Energy, Environmental and Information System. Semarang August 14-15, Vol 73 2018. EDP Sciences.
- [2] M. Kang, V. A. Lemieux, "decentralized identity-based blockchain solution for privacy-preserving licensing of individual-controlled data to prevent unauthorized secondary data usage", Ledger. Vol 6, pp: 126-151
- [3] A. Hakami, A. Al-Omary, "Secure Transaction Framework based on Encrypted One-time Password and Multi-factor", Proceedings of the International Conference on Data Analytics for Business and Industry, Bahrain, 25-26 October 2021, pp. 677-682.
- [4] R. Bernardini, "Cryptography - Recent Advances and Future Developments", IntechOpen, 2021.
- [5] A. Mittelbach, M. Fischlin, "The Theory of Hash Functions and Random Oracles : An Approach to Modern Cryptography", 1st Edition, Springer-Verlag Berlin Heidelberg 2021:
- [6] H. Mukhtar, "Kriptografi Untuk Keamanan Data" Edisi pertama, Deepublish, Yogyakarta, 2018.
- [7] R. Munir, "Kriptografi" Edisi Kedua", Bandung: Informatika, 2019.

- [8] J. Zhou et al. "Applied cryptography and network security workshops", Springer-Verlag Berlin Heidelberg, 2021.
- [9] R. Chaudhary et al. "Lattice-based public key cryptosystem for internet of things environment: Challenges and solutions", IEEE Internet of Things Journal, Vol. 6, No 3, 2018. pp. 4897-4909.
- [10] G. Mittal, S. Kumar, S. Kumar, "Novel public-key cryptosystems based on NTRU and algebraic structure of group rings". Journal of Information and Optimization Sciences, Vol. 42, No. 7, 2021. pp 1507-1521.
- [11] H. R. Yassein, A.A. Abidalzahra, N. M. Al-Saidi, "A new design of NTRU encryption with high security and performance", Proceedings of the 4th International Conference of Mathematical Sciences, Istanbul, Turkey, 17-21 June 2020, p. 080005
- [12] K. Daimi et al., "Computer and network security essentials", Springer Verlag, Berlin Heidelberg, 2018.
- [13] A. K. Sharma, S. Mittal. "Cryptography & network security hash function applications, attacks and advances: A review", Proceedings of the 3rd International Conference on Inventive Systems and Control, Coimbatore, India, 10-11 January 2019, pp. 177-188.
- [14] B. A. Forouzan, D. Mukhopadhyay, "Cryptography and network security", Mc Graw Hill Education (India) Private Limited New York, NY, USA, 2015.
- [15] S. Ghosh, S. Sampalli, "A survey of security in SCA-DA networks: Current issues and future challenges", IEEE Access, Vol 7, 2019, pp 135812-135831.
- [16] E. V. Waruwu, N. B. Nugroho, F. Sonata, "Penerapan Digital Signature Menggunakan Metode RSA Untuk Verifikasi Surat Keterangan Keaslian Ijazah SMA Swasta Bina Artha", Jurnal Cyber Tech, Vol. 1, No. 1, 2021. pp. 37-47.
- [17] Nuraeni, F., Y.H. Agustin, and I.M. Muharam, "Implementasi Tanda Tangan Digital Menggunakan RSA dan SHA-512 Pada Proses Legalisasi Ijazah", Proceedings of the Konferensi Nasional Sistem Informasi, Pangkalpinang, 8-9 March 2018, pp. 864-869.
- [18] B. Triand et al. "Digital Document Security on Legalize Higher Education Diplomas with Digital Signature and SHA-1 Algorithm", Proceedings of the 7th International Conference on Cyber and IT Service Management, 6-8 November 2019, pp. 1-5.
- [19] P. A. W. D. Putro, "Arumsari. Designing and Building Disposition-EL Application by Applying AES-256 and RSA-2048", Proceedings of the International Conference on Informatics, Multimedia, Cyber and Information System, 24-25 October 2019, pp. 163-168.
- [20] T. Yuniati, M. F. Sidiq, "Literature Review: Legalisasi Dokumen Elektronik Menggunakan Tanda Tangan Digital sebagai Alternatif Pengesahan Dokumen di Masa Pandemi". Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), Vol. 4 No. 6, 2020, pp. 1058-1069.
- [21] W. Pramusinto et al. "Implementation of AES-192 Cryptography and QR Code to Verify the Authenticity of Budi Luhur University Student Certificate", Jurnal Pendidikan Teknologi Kejuruan, Vol 3, No. 6, 2020, pp. 209-215.
- [22] J. Katz, Y. Lindell, "Introduction to modern cryptography", CRC press, 2020.
- [23] J. S. Kraft, L.C. Washington, "An introduction to number theory with cryptography", Chapman and Hall/CRC, 2018.
- [24] A. J. Menezes, P. C. Van Oorschot, S.A. Vanstone, "Handbook of applied cryptography", CRC press, 2018.
- [25] X. Yu, "Design of Aerospace QR Ticketing System Based on Mobile Devices", Proceedings of the 4th International Conference on Information Systems and Computer Aided Education, 24 September 2021, pp.2285-2287.
- [26] A. I. Chowdhury, M. S. Rahman, N. Sakib, "A study of multiple barcode detection from an image in business system", International Journal of Computer Applications, Vol. 181, No. 37, 2019, pp. 30-37.
- [27] Z. Azuan et al. "Mobile Advertising via Bluetooth and 2D Barcodes", Proceedings of the International Conference on Data Engineering 2015, Singapore, 10 August 2019. pp. 443-456.
- [28] E. Hari Charan, et al. "Electronic toll collection system using barcode technology in Nanoelectron-

- ics, Circuits and Communication Systems", Singapore, 2 August 2018. pp. 549-556.
- [29] D. D. Vo et al. "Barcode Image Restoration for Recognition of Product Information", European Journal of Engineering and Technology Research, Vol.4, No. 9, 2019. pp. 93-100.
- [30] R. Focardi, F. L. Luccio, H. A. Wahsheh, "Usable security for QR code", Journal of information security and applications, Vol. 48, No. 1, 2019, p. 102369.
- [31] R. Palomäki, "A distance-aware 2D barcode for mobile computing applications", Aalto University School of Science, Communication and Information Science, Finland, Master Thesis, 2018.
- [32] N. G. Kaziyeva, G. Kukharev, Y. Matveev. "Barcoding in biometrics and its development", Proceedings of the International Conference on Computer Vision and Graphics, Warsaw, Poland, 17 September 2018, pp. 464-471.
- [33] C. Guo, C.-C. Chang, S.-C. Chang, "A Secure and Efficient Mutual Authentication and Key Agreement Protocol with Smart Cards for Wireless Communications", International Journal of Network Security, Vol.20, No.2, 2018, pp. 323-331.
- [34] M. Qi, J. Chen, Y. Chen, A "secure authentication with key agreement scheme using ECC for satellite communication systems", International Journal of Satellite Communications and Networking, Vol. 37, No. 3, 2019, pp. 234-244.
- [35] M. Jiaqing, Z. Hu, H. Chen, W. Shen, "An Efficient and Provably Secure Anonymous User Authentication and Key Agreement for Mobile Cloud Computing", Wireless Communications and Mobile Computing, Vol. 2019, p. 4520685.
- [36] H. Loriya, A. Kulshreshta, D. Keraliya, "Security analysis of various public key cryptosystems for authentication and key agreement in wireless communication network", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, No. 2, 2017, pp. 267-274.
- [37] A. P. Siahaan, B. O. Elviwani, B. Oktaviana, "Comparative analysis of rsa and elgamal cryptographic public-key algorithms", Proceedings of the Joint Workshop KO2PI and The 1st International Conference on Advance & Scientific Innovation, 4 July 2018, pp. 162-171.

A comparative study of hash algorithms with the prospect of developing a CAN bus authentication technique

Original Scientific Paper

Asmae Zniti

Sidi Mohamed Ben Abdellah University,
Faculty of Sciences and Technologies (FST), Laboratory of Signals, Systems and Components (LSSC)
Route d'Imouzzer, Fez, Morocco
znitiasmae@gmail.com

Nabih El Ouazzani

Sidi Mohamed Ben Abdellah University,
Faculty of Sciences and Technologies (FST), Laboratory of Signals, Systems and Components (LSSC)
Route d'Imouzzer, Fez, Morocco
nabih.elouazzani@usmba.ac.ma

Abstract – In this paper, the performances of SHA-3 final round candidates along with new versions of other hash algorithms are analyzed and compared. An ARM-Cortex A9 microcontroller and a Spartan -3 FPGA circuit are involved in the study, with emphasis placed on the number of cycles and the authentication speed. These hash functions are implemented and tested resulting in a set of ranked algorithms in terms of the specified metrics. Taking into account the performances of the most efficient algorithms and the proposed hardware platform components, an authentication technique can be developed as a possible solution to the limitations and weaknesses of automotive CAN (Controlled Area Network) bus – based embedded systems in terms of security, privacy and integrity. From there, the main elements of such a potential structure are set forth.

Keywords: Hash algorithms, SHA-3, ARM-Cortex A9, FPGA, Number of Cycles, Authentication, CAN Bus

1. INTRODUCTION

In modern technology, embedded devices are smarter, more autonomous and better connected. Therefore, questions of information security are increasingly sensitive and have become of an utmost importance.

Hash functions are used for data integrity confirmation and as message authentication codes (MAC) or hash message authentication codes (HMAC). There exist several families of Security Hash Functions such as SHA-0, SHA-1, SHA-2, and SHA-3. In 1997, the National Security Agency NSA detected a major flaw in SHA-0 and so a new, improved algorithm –SHA-1 – was developed. This one, however, also suffered from severe cryptographic weaknesses and was later replaced by SHA-2 in 2002. Although as of yet, no significant cryptographic issue has been found in SHA-2, it was considered to be algorithmically too closely related to SHA-1. Since SHA-0, SHA-1 and SHA-2 suffer from these same limitations as well, we dismiss them, choosing instead to address the SHA-3 finalists [1] (Blake, Skein, JH, Grøstl, Keccak).

A number of studies have been conducted to compare the effectiveness of various hash algorithms. In 2013, R.K. Dahal et al. published a paper examining the performances of the SHA-3 finalists in addition to the widely used SHA-2 [2]. Their findings indicated that, among the SHA-3 finalists, Skein and Blake were the most effective while, according to digest length and block size, Grøstl, Keccak and JH followed.

In 2015, R. Sobti and Ganesan G. Geetha provided a performance evaluation of the SHA-3 finalists on ARM Cortex A8-based devices [3]. The results showed that Grøstl and JH were not efficient while Skein was a better option for long messages. They also found that Blake was a good option, as it outperformed Keccak for both 224 / 256- and 512-bit hash.

In 2018, a similar comparison was carried out on the ARM Cortex-M4 platform [4] with different results. Blake is the best choice as it performs better than all the algorithms for all message digests, regardless of the input size. Skein is also a good second option if, for higher security margins, we need a 512-bit hash instead of a 256-bit hash.

Numerous factors affect the performances of a hash algorithm. For example, Skein may be a better choice for long messages, while Blake may be better for short messages. It is important to consider the specific application when choosing the appropriate hash algorithm since no single algorithm is necessarily the best in all cases.

The proposed study provides a more comprehensive and up-to-date comparison of hash algorithms by implementing the SHA-3 finalists, as well as other newer common hashing functions, such as Blake2, Shake, Kangaroo Twelve and Blake3 on ARM cortex-A9 also taking their speed performance on an FPGA platform into consideration [5, 6]. As a result, the best algorithm is determined and the validity of the speed is verified.

As a potential application, an enhanced automotive CAN bus network [7] can be implemented based on a new structure relying on an authentication technique [8, 9, 10]. Considering the issue that a CAN bus lacks the security features such as message authentication and is therefore vulnerable to spoofing attacks [11, 12, 13], an effective solution may consist in implementing a hash process. Each message on the bus must be hashed with a key using a hash algorithm to form a message authentication code (MAC), thus allowing each node to check the authenticity of a received message. The process is mainly intended to generate two frames. The first deals with the transmission of data, while the second manages the authentication and filtering of unauthorized frames.

The structure of the paper is organized as follows. In Section 2, we briefly describe the SHA-3 Hash Function contenders. In Section 3, we present an overview of the CAN bus along with the principle of an authentication solution. Section 4 gives an outline of the methodology and the tools used for evaluation. Section 5 is dedicated to the hardware simulation results and performance analysis. The hash algorithm applied to produce digests is selected and the authentication time is calculated in section 6. Finally, Section 7 provides the conclusion.

2. AVAILABLE HASH ALGORITHMS

2.1. SECURE HASH ALGORITHM

A secure hash algorithm (SHA) is a standard invented by the National Institute of Standards and Technology (NIST) [14], based on the Message Digest (MD5) algorithm [15]. As the SHA-1 algorithm has already been cracked and SHA-2 proved to suffer from the same weaknesses, the NIST launched a public competition in November 2007 to make out a new cryptographic hash algorithm called SHA-3. In October 2012, NIST declared the Keccak algorithm as the winner of the SHA-3 competition.

2.2. THE SHA-3 FINALISTS

Keccak [16] was chosen from a range of five very strong candidates (Skein [17], Blake [18], Grøstl [19] and, JH [20]). NIST stated in its final report [21], that all five fi-

nalists had acceptable performances, and that any of the finalists would have represented an effective option for SHA3. In terms of performances, the report noted that some of the five algorithms operate well in software, while others appear more efficient in hardware.

2.3. THE BLAKE ALGORITHM FAMILY

Blake2 [22] is an improved version of Blake, provided in 2012 after Keccak was selected as SHA3. Blake2 was engineered to take full advantage of Blake's strengths and optimize it for modern applications. Blake2 comes in two main types: Blake2b which is optimized for 64-bit platforms and Blake2s for smaller architectures.

A successor of Blake 2, Blake 3 [23], created in 2020, was developed to be as fast as possible. The compression function of Blake3 is closely based on that of Blake2s.

2.4. THE KECCAK ALGORITHM FAMILY

Shake was declared by NIST in August 2015 as a part of the SHA-3 family. It combines two eXtensible Output Functions (XOFs), Shake128 and Shake256.

Kangaroo Twelve [24] is another XOF based on a reduced number of rounds (12 rounds) of the SHA-3 permutation function (Keccak [1600]). It is designed to be faster than SHA-3 and Shake while maintaining its flexibility and security.

3. CAN BUS OVERVIEW

3.1. CAN BUS PROTOCOL

The Controller Area Network (CAN) [7] is a serial communication bus that operates according to a specific standard for efficient and reliable information transmission between sensors, actuators, controllers, and other nodes in real-time applications.

On a CAN bus, the communication between different Electronic Control Units (ECUs) is achieved through four frames: data frame, remote frame, error frame, and overload frame.

As an example, Fig. 1 shows all the fields that make up the whole data frame. The data field length can reach up to 8 bytes, depending on the Data Length Code (DLC) word. A unique identifier is also assigned in order to manage both data transmission priorities between different nodes and filtering these upon reception. The size of the Identifier field is 11 bits for CAN version 2.0A and 29 bits for version 2.0B.

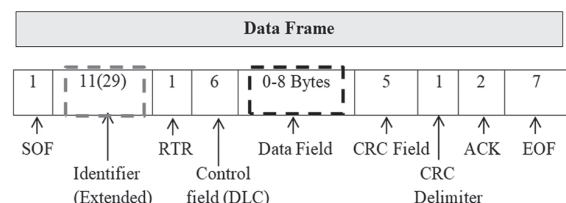


Fig. 1. CAN data frame structure

3.2. CAN BUS LIMITATIONS

As previously mentioned, the CAN bus is vulnerable to attacks perpetrated by malicious codes, leading to software damage and physical harm [25]. Amongst the weaknesses, we find:

- Non-confidentiality and transmission of unencrypted messages, which animate replay attacks and vehicle espionage.
- Absence of the authenticity and non-repudiation, which allows attackers to send arbitrary frames on the network or even transmit valid messages to trigger certain actions.
- Integrity: An attacker is able to add, delete, or modify any type of data carried by the relayed message.
- Availability: By sending high-priority messages, nodes are prevented from responding, which causes a denial of service (DoS), and consequently affects the system availability.

3.3. PRINCIPLE OF AN AUTHENTICATION TECHNIQUE

The fundamental idea depends on a system that includes a monitoring node made from an FPGA, equipped with a particular CAN controller, responsible for authenticating each message by verifying the MAC, generated by (1):

$$MAC = \text{hash function}(ID_i, D_i, FC_i, Key_i) \quad (1)$$

where ID_i is the CAN-ID (11 bits), D_i indicates the data of the message i (64 bits), FC_i represents a complete monotones counter for the message i of 32 bits, and

KEY_i denotes the encryption key for the node i encoded on 128, 256 or 512 bits.

Message data and the MAC are transmitted on two separate frames.

The planned protocol, illustrated in Fig. 2, will consist in computing the hash value by means of the CPU-based nodes participating in the communication, and performing the authentication of each CAN packet thanks to an FPGA-based monitoring node in charge of checking the hash value already calculated with its source.

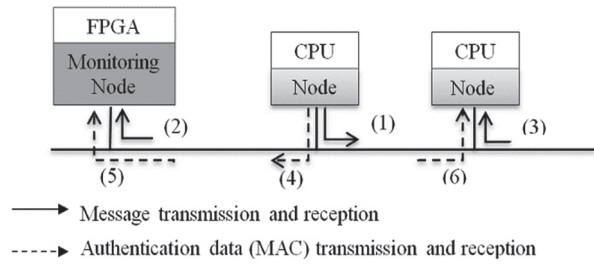


Fig. 2. Under-consideration communication protocol

4. ALGORITHM IMPLEMENTATION

4.1. PROCEDURE DETAILS

The flowchart in Fig. 3 shows the entire process of analyzing and comparing the performances in terms of the number of cycles and the execution time.

The algorithms of interest are run on an ARM Cortex A9-based platform, ranked according to their performances and compared to those obtained by means of an FPGA circuit as described in [5, 6].

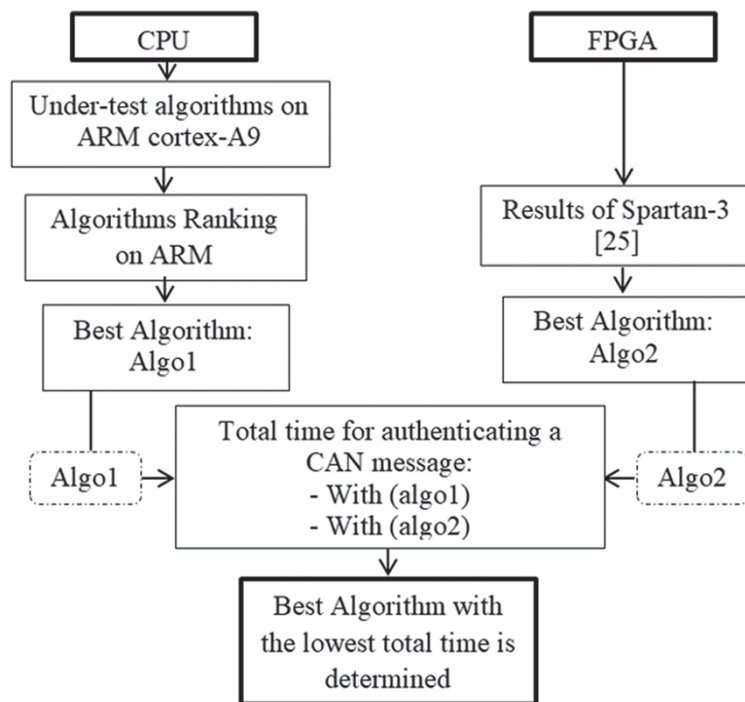


Fig. 3. Evaluation process of performance

4.2. ARM IMPLEMENTATION DATA

Knowing that 32-bit microcontrollers are widely used in embedded systems, particularly in the automotive industry, and in order to carry out simulations as closely as possible to real cases, the ARM Cortex A9 has been chosen as a testing tool. Three input sizes are considered depending on key bit lengths according to (1). The results are shown in Fig. 4.

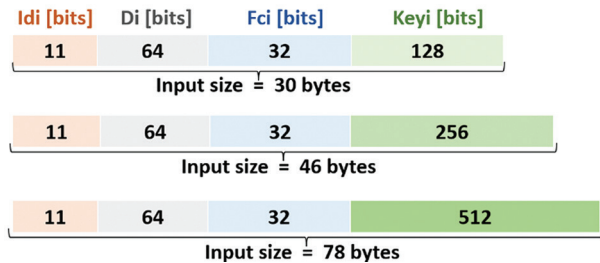


Fig. 4. Considered hash input sizes

In addition, running a hash algorithm on a Cortex A9-based platform requires a specific number of cycles regardless of the CPU frequency. As a result, the number of cycles is chosen as a metric for assessing performance.

5. RESULTS AND ANALYSIS

5.1. ARM CORTEX A9-BASED SIMULATIONS

Table 1 presents the cycles needed for 256 hash size of different algorithms taking three input lengths into account.

Table 1. Number of cycles for different input sizes

Algorithm	Number of Cycles [cycles]		
	Input size = 30 Bytes	Input size = 46 Bytes	Input size = 78 Bytes
Blake2s	13338	13180	21972
Blake	18866	18926	55376
Blake3	37002	69334	82278
Kangaroo Twelve	48320	83310	169190
Skein 256-256	64080	79646	79646
Skein 512-256	64152	64588	80190
JH	261690	261472	385978
Keccak	398660	398430	398976
Shake	558136	546286	547622
Grøstl	1682462	1690474	2789116

In this particular case of short inputs (less than 78 bytes), it is obvious that Blake2s and Blake outperform all the other contenders, providing the best choice.

Indeed, the results show that Blake2s beats all the other algorithms and presents the best speed of execution while Grøstl comes in a distant last place in comparison to other candidates.

Therefore, in real-case applications such as automotive networks, blake2s can be the core software code of any improved data transmission protocol.

5.2. FPGA-BASED RESULTS

As described in [5, 6], the FPGA Spartan-3 is a hardware platform that allows the considered hash algorithms to be run, highlighting major performance parameters from which a substantial advantage for the Keccak algorithm can be seen.

Table 2 mainly shows the results obtained through the FPGA of the two chosen algorithms according to the CPU and FPGA-based analysis.

Table 2. Performances of Keccak and Blake on Spartan-3 FPGA

Algorithm	Tclk [ns]	Bloc size	Rounds	Throughput [Mbps] = Blocsize / (Rounds*Tclk)
Keccak	9.75	1088	24	4650
Blake2s	43.4	512	10	1179.72

With regard to the FPGA implementation, Keccak has an advantage over Blake2s by providing the most suitable parameter values. Thus, Keccak is still considered the most appropriate solution.

6. COMPARISON OF BLAKE2S AND KECCAK

As a last step in the analysis and comparison procedure and from knowing the respective strength of Keccak and Blake2s, the focus in this section falls mainly on the entire processing time combining ARM and FPGA devices.

In the prospect of using an FPGA platform and CPU-based circuits in a protected CAN bus system, the processing time required by both components is computed by means of (2) and (3) respectively, with a CPU frequency of 667 MHz.

$$\text{Processing time (CPU)} = \text{Number of Cycle} / \text{Frequency} \quad (2)$$

$$\text{Processing time (FPGA)} = \text{Input Size} / \text{Throughput} \quad (3)$$

Table 3 shows the total time needed to generate an authentication message in this case.

Table 3. Keccak and Blake2S performances

Algorithm	Processing time [μs]					
	Input size = 30 Bytes		Input size = 46 Bytes		Input size = 78 Bytes	
	ARM	Sparatan3	ARM	Sparatan3	ARM	Sparatan3
Keccak	597.69	0.05	597.35	0.08	598.16	0.14
	Total = 597.74		Total = 597.42		Total = 598.30	
Blake2s	20.00	0.20	19.76	0.31	32.94	0.53
	Total = 20.20		Total = 20.07		Total = 33.47	

We notice, from Table 3, that the CPU processing contribution accounts for a large part of the total time for both Keccak and Blake2s. In addition, Blake2s offers a reduced message generating time ranging from 20 to 34 microseconds, which is 30 times faster than Keccak.

As a result, the use of the blake2s algorithm requires a maximum total hash time of 33.47 microseconds. Although the latency of a CAN message in automotive systems is typically around a few milliseconds, a few tens of microseconds is readily acceptable, as it would not have a significant effect on the CAN communication latency.

7. CONCLUSION

In this paper, an evaluation has been applied to the SHA-3 contenders' algorithms through an ARM Cortex A9 processor. A relevant comparison has also been performed involving an FPGA implementation to determine the fastest platform. From the standpoint of an ARM evaluation, the comparison shows a substantial win for Blake2s algorithm whereas Keccak offers excellent performances on FPGA circuits.

However, it is demonstrated that the CPU-based platform's impact on processing time is far more important than that of FPGA-based circuits. The comparison between Keccak and Blake2s shows that the latter is more likely to suit perfectly the targeted performances of the planned security system.

The use of cryptographic algorithms in automotive CAN bus systems can introduce limitations of computing speed related especially to the time response when dealing with real-time demands. However, considerable room for improvement with respect to security and data protection can be achieved thanks to the possibility of including a monitoring node along with the application of the justifiably chosen Blake2s that can contribute to a greater effectiveness.

This process is meant to be part of an authentication system within the CAN bus aiming to overcome the vulnerability problems of the network. As a matter of fact, implementing a prototype and running hardware simulations according to the guidelines of the enhanced network will be the follow-up task with respect to the development of the ongoing process.

8. REFERENCES:

- [1] NIST, SHA-3 Competition (2007-2012), <https://csrc.nist.gov/groups/ST/hash/sha-3/> (accessed: 2017)
- [2] R. Dahal, J. Bhatta, T. Dhamala, "Performance Analysis of Sha-2 and Sha-3 Finalists", *International Journal on Cryptography and Information Security*, Vol. 3, No. 3, 2013, pp. 1-10.
- [3] R. Sobti, G. Geetha, "Performance Comparison of Keccak, Skein, Grøstl, Blake and JH: SHA-3 Final Round Candidate Algorithms on ARM Cortex A8 Processor", *International Journal of Security and Its Applications*, Vol. 9, No. 12, 2015, pp. 367-384.
- [4] R. Sobti, G. Ganesan, "Performance Evaluation of SHA-3 Final Round Candidate Algorithms on ARM Cortex-M4 Processor", *International Journal of Information Security and Privacy*, Vol. 12, No. 1, 2018, pp. 63-73.
- [5] J. Sugier, "Improving FPGA implementations of BLAKE and BLAKE2 algorithms with memory resources", *Proceedings of the 12th International Conference on Dependability and Complex Systems*, Brunów, Poland, 2-6 July 2017, pp. 394-406.
- [6] J. Sugier, "Low cost FPGA devices in high speed implementations of KECCAK-f hash algorithm", *Proceedings of the 9th International on Dependability and Complex Systems*, Brunów, Poland, 30 June - 4 July 2014, pp. 433-441.
- [7] R. Bosch, "Can specification version 2.0", Postfach, Stuttgart, Germany, Technical Report Bosch, 1991.
- [8] O. Avatefipour, A. Hafeez, M. Tayyab, H. Malik, "Linking received packet to the transmitter through physical-fingerprinting of controller area network", *Proceedings of the IEEE Workshop on Information Forensics and Security*, Rennes, France, 4-7 December 2017, pp. 1-6.
- [9] P. Mundhenk, A. Paverd, A. Mrowca, S. Steinhorst, M. Lukasiewicz, S. A. Fahmy, S. Chakraborty, "Security in automotive networks: Lightweight authentication and authorization", *ACM Transactions on Design Automation of Electronic Systems*, Vol. 22, No. 2, 2017, pp. 1-27.
- [10] J. Van Bulck, J. T. Mühlberg, F. Piessens, "VulCAN: Efficient component authentication and software isolation for automotive control networks", *Proceedings of the 33rd Annual Computer Security Applications Conference*, Orlando, USA, 4-8 December 2017, pp. 225-237.
- [11] H. Zhang, X. Meng, X. Zhang, Z. Liu, "CANsec: A Practical in-Vehicle Controller Area Network Security Evaluation Tool", *Sensors*, Vol. 20, No. 17, 2020, p. 4900.

- [12] A. Zniti, N. El Ouazzani, "Implementation of a blue-tooth attack on controller area network", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 21, No. 1, 2021, pp. 321-327.
- [13] Y. Yang, Z. Duan, M. Tehranipoor, "Identify a spoofing attack on an in-vehicle can bus based on the deep features of an ecu fingerprint signal", *Smart Cities*, Vol. 3, No. 1, 2020, pp. 17-30.
- [14] M. J. Dworkin, "SHA-3 standard: Permutation-based hash and extendable-output functions", NIST, Gaithersburg, MD, USA, Technical Report NIST FIPS-202, 2015.
- [15] R. Rivest, "The MD5 message-digest algorithm", IETF Network Working Group, MA, USA, Technical Report RFC 1321, 1992.
- [16] G. Bertoni, J. Daemen, M. Peeters, G. V. Assche, "The Keccak sha-3 submission", NIST SHA-3 Competition (Round 3), Gaithersburg, MD, USA, Technical Report 03, 2011.
- [17] N. Ferguson, S. Lucks, B. Schneier, D. Whiting, M. Bellare, T. Kohno, J. Callas, J. Walker, "The skein hash function family", NIST SHA-3 Competition (Round 3), Gaithersburg, MD, USA, Technical Report 1.3, 2010.
- [18] J. P. Aumasson, W. Meier, R. C. Phan, L. Henzen. "The Hash Function BLAKE", 1st Edition, Springer Berlin Heidelberg, 2014.
- [19] P. Gauravaram, L. R. Knudsen, K. Matusiewicz, F. Mendel, C. chberger, M. Schlfifer, S. S. Thomsen, "SHA-3 proposal grostel", NIST SHA-3 Competition (Round 3). Gaithersburg, MD, USA, Technical Report 2.0.1, 2008.
- [20] H. J. Wu, "The hash function jh", NIST SHA-3 Competition (Round 3), Gaithersburg, MD, USA, Technical Report 42, 2011.
- [21] NIST: Third-round report of the SHA-3 cryptographic hash algorithm competition, <http://www.nist.gov/hash-competition> (accessed: 2012)
- [22] J. P. Aumasson, S. Neves, Z. W. O'Hearn, C. Winnerlein, "BLAKE2: Simpler, smaller, fast as MD5", Proceedings of the 11th International Conference on Applied Cryptography and Network Security, Banff, AB, Canada, 25-28 June 2013, pp. 119-135.
- [23] S. Neves, J. O'Connor, J.P. Aumasson, Z. Wilcox-O'Hearn, BLAKE3: One function, fast everywhere, <https://github.com/BLAKE3-team/BLAKE3-specs/blob/master/blake3.pdf> (accessed: 2020)
- [24] G. Bertoni, J. Daemen, M. Peeters, G. V. Assche, R.V. Keer, B. Viguier, "KangarooTwelve: Fast hashing based on Keccak-p", Proceedings of the 16th International Conference on Applied Cryptography and Network Security, Belgium, 2-4 July 2018, pp. 400-418.
- [25] L. Pan, X. Zheng, H. Chen, T. Luan, H. Bootwala, L. Batten, "Cyber security attacks to modern vehicular systems", *Journal of Information Security and Applications*, Vol. 36, 2017, pp. 90-100.

Towards Auto Contract Generation and Ensemble-based Smart Contract Vulnerability Detection

Original Scientific Paper

K. Lakshminarayana

Puducherry Technological University,
Ph.D Scholar, Department of Computer Science and Engineering,
Puducherry, India
kodavali.lakshmi@pec.edu

K. Sathiyamurthy

Puducherry Technological University,
Faculty of Computer Science and Engineering,
Puducherry, India
sathiyamurthyk@ptuniv.edu.in

Abstract – Smart contracts (SC) are computer programs that are major components of Blockchain. The "intelligent contract" is made up of the rules accepted by the parties concerned. When the transactions started by the parties obey these established rules, then only their transactions will be completed without the involvement of a third party. Because of the simplicity and succinct nature of the solidity language, most smart contracts are written in this language. Smart contracts have two limitations, which are vulnerabilities in SC and that smart contracts can't be understood by all stakeholders, especially non-technical people who are involved in the business, since they are written in a programming language. Hence, the proposed paper used the XGBoost model and BPMN (Business Process Modeling Notation) tool to solve the first and second limitations of the SC respectively. Attackers are drawn to attention because of the popularity and fragility of the Solidity language. Once smart contracts have been launched, they can't be changed. If that smart contract is vulnerable, attackers may then cash it. BPMN is used to represent business rules or contracts in graphical notation, so everyone involved in the business can understand the business rules. This BPMN diagram can be converted into a smart contract template through the BPMN-SOL tool. A few publications and existing tools exist on smart contract vulnerability detection, but they require more time to forecast and interpretation of vulnerability causes is also difficult. Thus, the proposed model experimented with several deep learning approaches and improved F1 score results by an average of 2% using the XGBoost model based on the ensemble technique to detect vulnerabilities of SCs, which are: Denial of Service (DOS), Unchecked external call, Re-entrancy, and Origin of Transaction. This paper also combined two important features to construct a data set, which are code snippets and n-grams.

Keywords: Blockchain, Smart Contract Vulnerabilities, Ethereum, Machine Learning, Ensemble Model, BPMN

1. INTRODUCTION

A Blockchain works in a decentralized environment and it has a sequence of blocks that are connected using cryptographic techniques [1]. As shown in Figure 1, each block consists of transaction data, a hash of the preceding block, and a timestamp. By its design, blockchain is unsusceptible to changing data by its design. In a Blockchain, transactions between two parties are recorded in an efficient, verifiable, and permanent way [2]. Such a Blockchain can present an innovative solution to long-standing problems of security related to data storage in centralized systems. Blockchain can be considered the new face of cloud computing and

is expected to reshape organizational and individual behavioral models.

An important feature of a Blockchain is, that it is a distributed database. It means no centralized database or server exists. Instead, the same Blockchain is duplicated on every node of the network. Each node in the system receives a duplicate of Blockchain where all chunks have a grade of dealings in an encrypted format using asymmetric keys. Due to the complexity of mathematical formulas used in cryptography techniques, it is practically impossible to guess the keys and crack the transactions. The sender can use his private key to encode a message to be sent, and the recipient can use his pub-

lic key to decrypt the message. Every new transaction is broadcast and updated to all the network nodes to maintain a consistent database across the whole Blockchain network [3]. Bitcoins are the major general Blockchain stand in the world. Ethereum is another popular Blockchain that introduces smart contracts.

Smart Contracts (SC) are the programs for predefined rules which are deployed into the Blockchain and these programs execute automatically to make sure that every transaction has to satisfy the predefined conditions to complete the transaction. Smart Contracts work based on simple conditional statements. Smart Contracts are playing a more vital role in business among a group of untrusted people, where every transaction can be completed according to rules agreed by all business stakeholders without the involvement of third-party verification [4]. Initially, SC basis codes are written in a high-level language, for example, Solidity

by designers. The source code is compiled into byte codes (EVM code) by a compiler, it is in a hexadecimal arrangement. These byte codes can be converted into EVM instructions and are called opcodes [5].

Broadly three reasons attackers are focusing on smart contracts: first, the smart contracts of Ethereum are mainly money oriented transactions; secondly, after being deployed into the Blockchain, it is not possible to alter vulnerable SC; and finally, smart contracts have no defined measures to determine the quality of smart contracts [6]. Many smart contract assaults in 2016 led to large money losses (multi-million dollar losses) as a result of vulnerabilities in SCs. In smart contracts on Ethereum, it's currently worth focusing on automated deep study models to effectively detect SC vulnerabilities [7], especially in financial matters such as money transfers and more complicated code.

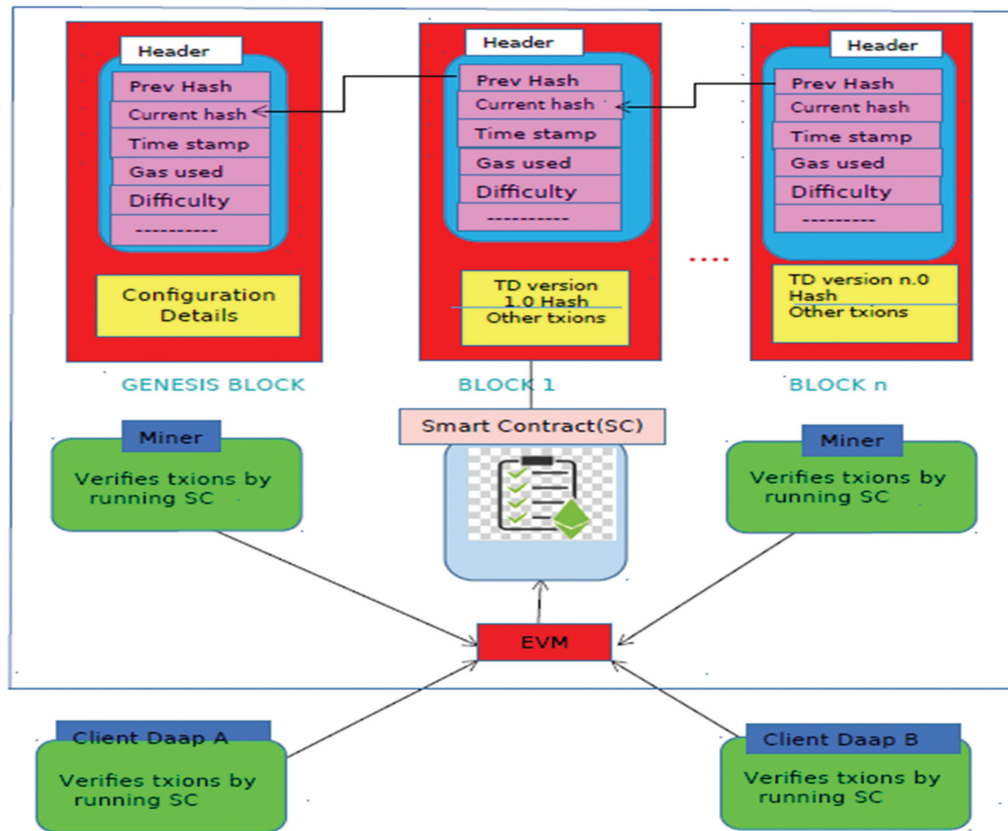


Fig 1. Ethereum Blockchain Network

Smart contract vulnerabilities are divided into four categories, 1. security, 2. functional, 3. developmental, and 4. operational [11]. Security concerns include re-entrancy, external contract, DOS - denial of service, the use of tx.origin, unchecked external call, and usage of send() in place of transfer(). Functional vulnerabilities are locked money, integer division, integer underflow, integer overflow, unsafe interface type, and reliance on time stamps. Development concerns involve infringement of token API, private modifier, non-compiler version fixation, violation of the style guide, duplicated back func-

tion, and degree of implied visibility. Finally, operational problems include byte array and expensive loop vulnerabilities. The major focus of this article is on security concerns that consist of unchecked external calls, re-entrancy, DOS, and the origin of the transaction.

Reentrancy vulnerability[10,15]: In which they can take on the control flow and modify your data that is not expected by the call function. There are numerous shapes this bug class might take. The initial release of this problem is single-function reentrancy in which

functions can be called repeatedly before the original call is made. This might lead to harmful interactions between the multiple calls of the function. The other release of this problem is the re-entrancy of a cross-function, in which an attacker can also attack the same state with two functions [8]. As re-entrancy can occur through several functions and even between different contracts, a single function will not be enough to prevent re-entrancy. Instead, all internal work (i.e. state modifications) has been recommended first, and then the external function has to be called to prevent re-entrancy vulnerability.

DOS vulnerability: A specific quantity of gas (transaction fee) is required for the execution of smart contract functions. The Ethereum network establishes a gas limit for every block and should not exceed that amount of all transactions in a block. Programmable statements in smart contracts that cause DOS (Denial of Service), if the gas limit is exceeded when these statements are executed. DOS vulnerability may be caused in scenarios such as 1) A loop variable with a value higher than or less than 382 depends more or less on the network gas limit. 2) Work with unfamiliar array sizes.

Vulnerability of transaction origin: The keyword called "tx.origin" in solidity language indicates the address of the account that began a transaction. For example, consider the sequence of call series X--> Y and Y--> Z. From the Z viewpoint, msg.sender is Y, and tx.origin is X. The "tx.origin" keyword can sometimes lead to dubiety, therefore try to avoid the use of "tx.origin" for authorization. Instead, it can be handled with msg.sender.

Machine Learning (ML) algorithms are classified by learning style, consisting of supervised, semi-supervised, and unsupervised. In supervised learning, input or training data has a predefined label. Initially, a classifier has to be designed with appropriate layers to train on training data and to predict the label of test data. The classifier has to be tuned well to get a good level of prediction accuracy. In unsupervised learning, training data does not have a label, hence the classifier is designed to cluster unsorted data based on similarities and variance.

Wang Wei et al [5] presented an automated vulnerability detection model for smart contracts using the XGBoost machine learning model by extracting bigram features from SC opcodes. The limitation of this paper is that bigrams (2-gram) may not always be suitable to detect all types of vulnerabilities, because some vulnerabilities may require more than bigram features. Interpretation of opcodes is more difficult compared with high-level source code. The proposed framework uses an ensemble-based XGBoost supervised model [5] to perform multi-label classification to detect SC vulnerabilities with the help of a data set created by n-gram features which are extracted from high-level SC source code. Ensemble algorithms join the outcomes of a set of simple and feeble models to get better predictions

than those that are obtained using a single learning algorithm. The association of this remaining paper is organized as follows: Part 2 gives literature work on smart contract vulnerability detection systems using machine learning; part 3 explains the proposed work for vulnerability detection using the XGBoost learning model; part 4 demonstrates experiment details and comparison outcomes; and finally, conclusion and future scope will be in part 5.

2. LITERATURE WORK

Recently, a few papers [5,6,9,10,11] were published on behalf of smart contract susceptibility detection using different ML techniques. Jian-Wei Liao et al. [6] present smart contract susceptibility detection using machine learning and fuzz testing techniques. The Authors used existing SC vulnerability detection tools, which are Oyente and Remix, to label training data sets. These static detection tools are more time-consuming [12]. The Authors also stated that to detect vulnerabilities of SC, it requires SC skilled people or predefined patterns of vulnerabilities. They extracted features from SC opcodes to prepare the data set.

Pouyan Momeni et al [9] presented smart contract security analysis with machine learning techniques, and the authors used existing traditional tools, which are Mythril [13] and Slither to label the SCs that are present in the dataset. These SC are given to a solidity parser as an input, and it generates an AST (Abstract Syntax Tree). Processing of AST is straightforward and easy to interpret. Features can be extracted comfortably from this AST as its output. The traditional tools used in the paper have been taking more time to predict vulnerabilities [12]. Feng Mi et al. [14] presented a paper on the automatic detection of SC vulnerabilities using deep learning. The authors prepared a data set with extracted features from the SC byte code. Moreover, the interpretation of byte code makes it difficult to analyze results. Peng Qian et al. [7] proposed graph neural networks for smart contract vulnerability detection with the help of expert knowledge. The authors constructed a graph for the extracted patterns from a given SC. Authors developed an open-source tool to extract patterns or features. Hence, in the proposed work, the open source feature extraction tool has been used and tailored as per the proposed work requirements.

Lakshminarayana. K et al [15] experimented with basic classification methods, which are binary classification, multiclass classification, multi-label classification, and auto encoding techniques to detect smart contract vulnerabilities, which are reentrancy, DOS, and Tx.origin. The proposed paper also tries to improve detection results of the same vulnerabilities using a combination of two techniques, which are the contract snippets, n-gram features, and the XGBoost classification technique. Yiping Liu et al. [19] presented an SC vulnerability detection model based on symbolic execution by taking SC assembly code (opcode) as input to generate a control

flow graph. Zhang L et al [20] presented an SC vulnerability detection model based on an information graph and ensemble technique. Input for this model is considering SC opcodes to find the critical opcode sequence for each vulnerability. But the proposed paper has been using high-level source code directly as input since SC source code is easier to trace the source of vulnerability than SC byte code or SC opcodes. Nowadays, deep learning techniques have been extensively used in many areas in real life, for example, to prevent COVID by tracing social distances [16], network behavior monitoring [17], to help programmers who feel it is difficult to learn by identifying their mistakes and suggesting corrections [18], and also for the detection of unusual activities in the health care sector, etc.

Zhenguang Liu et al. [10] present automated re-entrancy detection for SC. This paper used BLSTM for the classification task. The authors propose contract snippets (keywords) to capture semantic information from a given SC. The limitation of this paper is that it considers every word of each line in an SC to prepare a feature set. This may increase the number of features, but it may lead to a reduction in classification accuracy. Wesley Joon-Wie Tann et al [11] presented the LSTM machine learning model for safer smart contracts. LSTM will consider a sequence of opcode features to detect SC vulnerabilities. But LSTM doesn't care about the inspection of data and control-flow possessions (ex: loops and function calls). K. Frantzet al[21] and Luciono B et al [22] proposed methods to convert BPMN diagrams to solidity smart contract templates. Authors

in [21] developed a domain-specific language (DSL) called ADICO-Solidity DSL for conversion from BPMN to solidity code. Authors in [22] prepare a Petri net graph as an intermediate step from BNMN elements and then convert them to solidity code from Petri nets. Each of the above papers has its pros and cons. Hence, the proposed paper combines the pros of the above papers, which are the usage of a parser, contract snippets, and usage of n-grams to prepare the data set in the proposed SC vulnerability system, and also does work towards auto SC generation using the BPMN-SOL compiler [23,24].

3. PROPOSED SYSTEM

The general building of the proposed model is shown in Figs. 2(a) and 2(b). In the proposed system, a smart contract is supplied as input to the solidity parser. It performs preprocessing steps like removing comments and identifying functions and loops. From the parser output, code snippets (important keywords) and n-grams can be extracted as shown in Table1.

These n-grams play a major role in identifying vulnerabilities. Now it is possible to prepare a data set by considering n-grams as features, and they could be labeled according to the n-grams found in the SC. As shown in figure 3, different classes of vulnerability (C1 to C4) are correlated with different disjoint sets of n-gram features (N1 to N7). It applies to all solidity smart contracts if we provide high-level solidity source code as an input instead of byte code or opcode of smart contracts.

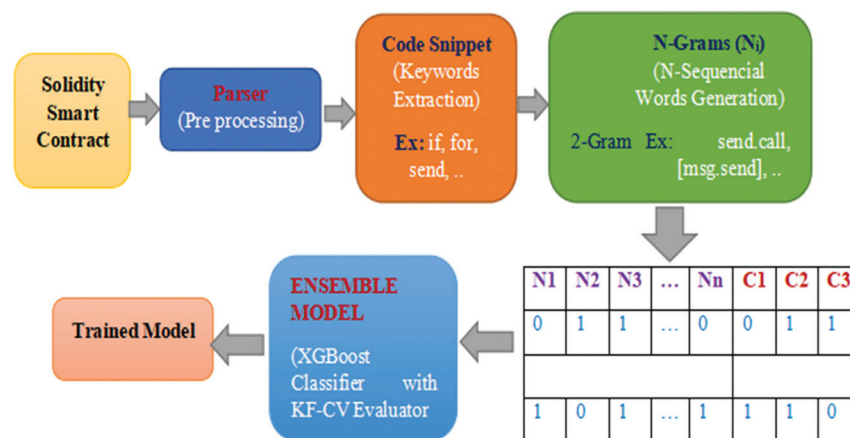


Fig 2(a). Training a Machine Learning Model

Table 1. Examples for Code Snippets and n-grams from smart contracts

Code Snippets (Key Words)	n-grains
if, (), msg, sender,	msg. sen der.ca 11.va lue, a ddr.transfer(*)
call, value, ., [,],	msg.sender, tx.origin()9 solidity ^9
While, transfer, function,	funtion(), [msg.sender]=0
return. require. revert.	msg.sender.transfer, addr.send(""),
addr,	if(*addr.send(*) revert, if(*(),

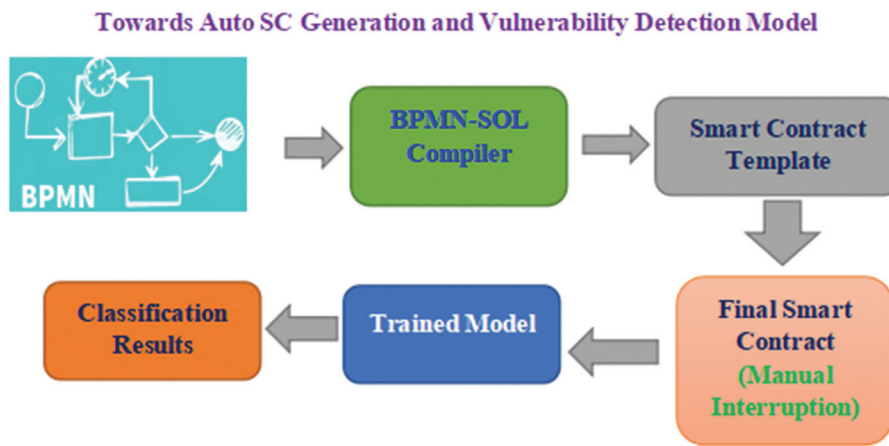


Fig 2(b). Smart Contract Template generation and testing a Model

	N-Gram Features							Vulnerability Classification			
	N1	N2	N3	N4	N5	N6	N7	C1	C2	C3	C4
1	1	0	0	0	0	1	1	1	0	0	0
2	0	1	0	0	0	0	0	0	1	0	0
3	0	0	1	1	0	0	0	0	0	1	0
4	0	0	0	0	1	0	0	0	0	0	1
5	1	0	0	0	1	1	1	1	0	0	1
6	0	0	1	1	1	0	0	0	0	1	1
7	0	1	1	1	1	0	0	0	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1

C1⇒ DOS Attack: **N1**: If(*0), **N6**: if(*LV*), **N7**: while(*LV*) [LV- Large Value]
C2⇒ Unchecked External Call: **N2**: addr.send()
C3⇒ Re-entrancy: **N3**: msg.sender.call.value, **N4**: [msg.sender]=0, [**N3,N4**],
C4⇒ Tx.origin: **N5**: Tx.origin

Fig 3. Data Set for multi-label classification to detect vulnerabilities

The data set is created after analyzing the smart contract source codes, which are collected from [25, 27] as per the algorithm shown in Figure4. In the literature on SC, many papers prepared data sets from the byte code of SC, which are difficult to analyze and difficult to confirm whether SC is vulnerable or not. Hence, this paper prepared datasets from high-level SC source code instead of from byte code or opcodes of SC.

Multi-label classification task: this requires one or more labels for each input sample as an output, and the outputs are required at the same time. The hypothesis is that the output labels depend on the inputs. In the proposed system, four classes of vulnerabilities, which are DOS (C1), unconditional external call (C2), reentrancy (C3), and transaction origin (C4), can be predicted.

This prepared data set is given to ensemble-based XGBoost classifiers to train the network, where it uses internal K-Fold cross-validation (KF-CV) to test and improve its training performance. Cross-validation means

an evaluation of machine learning models on a small sample of data. Cross-validation is mostly utilized to evaluate the skills of a machine learning model on invisible data in applied machine learning. A K-Fold CV is a collection of K sections/folds where each fold is utilized as a test set at a certain moment. Consider the case of a 5-fold CV, where K=5. In this case, the total data set is divided into five equal partitions. First, the first partition acts as a test set and the remaining partitions act as training sets. In the next iteration, the 2nd partition acted as the testing set and the remaining partitions acted as training sets. This process will continue for all five partitions, and finally, the results of all folds are averaged to predict the model's final performance.

XGBoost stands for Extreme Gradient Boosting. It is a scalable and tree-based boosting machine learning model. It is a popular and efficient open-source implementation. Gradient boosting is a learning method that tries to forecast a target variable by integrating estimates for several weaker and simpler models. In

Ensemble Learning, XGBoost is included in the boosting methods group. Ensemble learning consists of a group of predictors that offer higher prediction accuracy through several models. In gradient-boosted algorithms, the loss function is optimized, unlike in other booster techniques where incorrectly categorized branch weights are raised. XGBoost is a sophisticated gradient booster implementation with certain regulatory features. XGBoost features include, that it can be executed both on single and distributed systems (Hadoop, Spark) (regression and classification problems), parallel processing support, optimization of cache, and effective memory management for big data sets over RAM. It contains many regularizations that help to reduce overfitting problems. Auto tree pruning means

the decision tree does not increase further internally beyond specified limitations, can manage any missing information, has cross-validation integrated, and takes care of some outliers.

As illustrated in Figure 5 in the XGboosting, mistakes caused by earlier models may be rectified by successive models. The trees will be created in sequence to minimize the mistakes of the previous tree in every successive tree. The previous tree lists each tree and updates the remaining bugs. The successive trees in the series will thus find information from an updated residual version. Gradient improvement is only a framework into which any model may be plugged, but tree-based models offer superior results.

```

-----
Algorithm: Dataset_Preparation(contract.sol)
-----
Input      : Smart_Contract_Program (contract.sol)
Output     : Generate ngram_Vector[N1, N2, N3, N4, N5, N6, N7]

//ngrams_initialization
n1=if(*) // here * indicates any character
n2=addr.send(), n3=msg.sender.call.value, n4=[msg.sender]=0
n5=tx.origin, n6=if(*LV*) //if condition with large value(>382)
n7=while(*LV*) //while condition with large value(>382)

Function Feature_Extraction( Path/contract.sol )
  Remove comments in contract.sol file using solidity parser
  //split contract program into list of words using space delimiter
  cwords=contract.split()
  //Initialize ngram vector with all zeros
  [N1, N2, N3, N4, N5, N6, N7, C1, C2, C3, C4]=0
  if n1,n6,n7 present in cwords
    set N1, N6, N7, C1 with 1 //DOS_Vulnerability
  if n2 present in cwords
    set N2, C2 with 1 //Unchecked_External_Call Vulnerability
  if n3, n4 present in cwords
    set N3,N4,C3 with 1 //Re-entrancy Vulnerability
  if n5 present in cwords
    set N5,C4 with 1 //Tx.origin Vulnerability
  add ngram_vector[N1,N2,N3,N4,N5,N6,N7,C1,C2,C3,C4] into
  dataset as one record
end function
//Repeat above function for every smart contract, to prepare final data set.
-----

```

Fig. 4. Algorithm to generate data set from given smart contracts

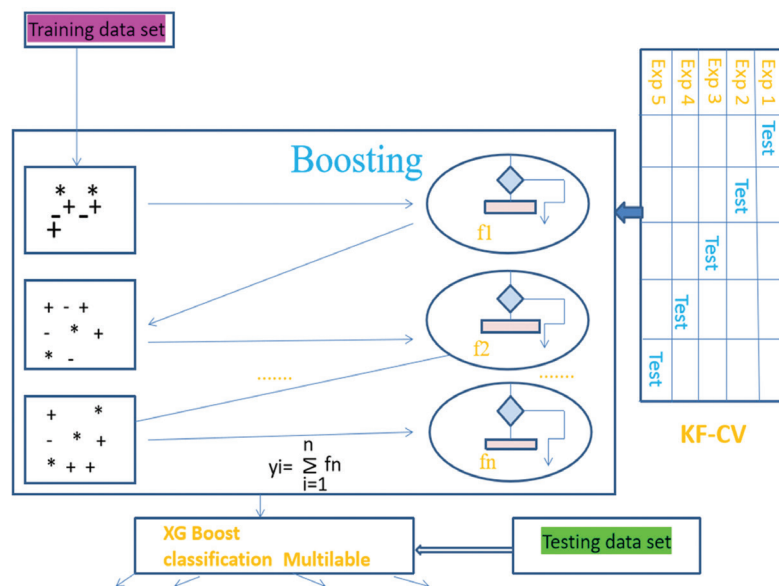


Fig. 5. XGBoost Classification Model

Once a machine learning model gets trained well, then it will classify or predict the smart contract vulnerabilities if we supply any SC as input for it. As smart contracts are written in programming languages, all stakeholders of a business can't interpret them to make them conform to whether they meet all business requirements or not. Hence, this paper used the Business Process Modeling Notation (BPMN) tools [25,26] to construct a graphical representation of a smart contract, as these BPMN diagrams are easy to understand by everyone, even if they do not have any technical knowledge, as shown in Figure 6. It includes functions

for withdrawal, deposit, and balance inquiry. This BPMN diagram can be converted into a smart contract template by the BPMN-SOL tool. It is a compiler to convert a BPMN file to a solidity file [23,24]. The BPMN-SOL tool internally follows the caterpillar engine to convert a given smart contract into a solidity code template [18]. The output of the BPMN-SOL tool is the SC template as shown in Figure 7 (sample code, it is not generated, we are still working on it). The SC template can't be processed by the solidity compiler. Hence, developer involvement is required to make slight changes to convert it into a final smart contract as shown in Figure 8.

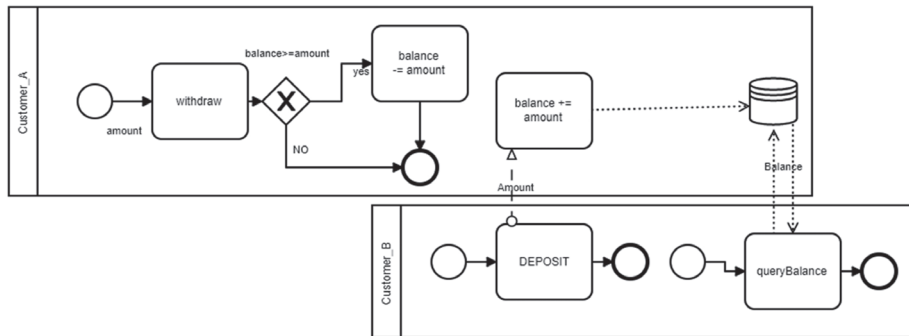


Fig. 6. BPMN diagram for Withdraw, Deposit, and Balance Enquiry functions

```

1  pragma solidity -version-;
2
3  contract ContractName {
4      mapping ( DataType => DataType) public balance;
5
6      function donate(address Customer_A) payable public{
7          balance[Customer_A] += amount;
8      }
9
10     function withdraw(DataType amount) public{
11         if (balance[Customer_A]>= amount) {
12             require(Customer_A.call.value(amount)());
13             balance[Customer_A]-=amount;
14         }
15     }
16
17     function queryBalance(DataType Customer_B) view public returns(DataType){
18         return balance[Customer_B];
19     }
20 }

```

Fig. 7. Sample Smart Contract Template for the given BPMN diagram (Fig. 6)

```

1  pragma solidity 0.4.24;
2
3  contract SimpleContract {
4      mapping( address => uint ) public balance;
5
6      function donate( address to ) payable public{
7          balance[ to ] += msg.value;
8      }
9
10     function withdraw( uint amount ) public{
11         if (balance[msg.sender] >= amount) {
12             require(msg.sender.call.value( amount )( ) );
13             balance[msg.sender ] - = amount;
14         }
15     }
16
17     function queryCredit( address to ) view public returns(uint){
18         return balance[ to ];
19     }
20 }

```

Fig. 8. Final Smart Contract after modifications to SC Template

This paper concentrated on eXtreme Gradient Boosting (XGBoost) to build a suggested system to discover vulnerabilities in smart contracts based on code snippets and n-grams. For comparison purposes, this paper also used Random Forest (RF), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) machine learning models.

Random Forest is a flexible and easy-to-use algorithm for classification tasks. It is a kind of supervised learning algorithm. Random forests may construct many trees from randomly selected data samples. Its final classification prediction result is based on the majority voting results of trees constructed by it. The robustness of this model depends upon the number of trees constructed by it. Even though the random forest is a good model, compared with gradient-boosted trees, it has lower accuracy. SVM is the extensively utilized classification process, and it aims to identify a hyperplane in positive or negative samples in each of the segments such that there is the highest margin for the two segments, where the classification system is very reliable and generalizes new samples. The KNN classification algorithm is also used frequently. It is efficient and straightforward. For the given test sample, k-samples nearest to the test sample are determined based on a certain distance measure, and then information from k-neighbors is predicted. In k-samples, the most common category marks are chosen as the prediction outcomes, depending on the majority voting.

4. EXPERIMENT DETAILS AND RESULTS

The Keras Python package is being used to experiment with deep-learning models. Keras is a user-friendly, free, and useful framework to create and evaluate deep learning models with a few code lines. The COLAB resource is helpful for executing all the Python programs needed for this task online without cost. First of all, we must upload the data sets to the Google Cloud (colab) before the applications are run.

Metrics[29]: The matrices used in the proposed work are confusion matrix, recall, precision, F1 score, Micro-F1, and Macro-F1 to measure the deep learning models (XGBoost, SVM, RF, and KNN) performance for multi-label classifications. The metric types used will direct us to choose better machine learning models. To assess the performance of multi-label classification, the useful measures are Micro-F1 and Macro-F1. Micro-F1 and Macro-F1 are called Global-F1 and Average-F1, respectively. These measures can be calculated from the confusion matrix. It involves variables TN, FN, FP, and TP, which stand for True Negatives, False Negatives, False Positives, and True Positives, respectively.

$$\text{Precision} = TP / (FP + TP);$$

$$\text{Recall} = TP / (FN + TP);$$

$$\text{F1-SCORE} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Both the Micro F1-score and the Macro F1-score are used to evaluate the performance of multi-label binary problems. For both, the best value is 1 and the worst

value is 0. The Micro-F1_score is defined as the harmonic mean of precision and recall. The Micro-F1_score measures the aggregated F1_score of all classes. Note that precision and recall have the same relative contribution to the F1_score. It can give high values even if the model is performing poorly on the rare labels since it gives more importance to the frequent labels. For the calculation of the micro-averaging F1-score, initially compute the sum of all false positives, true positives, and false negatives over all the labels. Then calculate the global precision and global recall from these sums. Finally, compute the harmonic mean to obtain the micro F1-score.

$$\text{Micro-F1_Score} = (2 * \text{Global Precision} * \text{Global Recall}) / (\text{Global Precision} + \text{Global Recall})$$

Macro-F1_score will treat all classes equally. It will give a low value for the models that only do well in the frequent classes while showing unsatisfactory results in the rare classes. Macro F1-averaging is calculated by computing the F1-score for each class and then averaging them.

$$\text{Macro-F1_Score} = (2 * \text{Average Precision} * \text{Average Recall}) / (\text{Average Precision} + \text{Average Recall})$$

For the XGBoost classifier, all metrics are calculated as shown in Table 2. The dataset consists of a total of 1380 records, with 320, 330, 350, and 380 records for C1, C2, C3, and C4 respectively. The detailed calculation of Micro and Macro F1-scores of XGBoost classifier is shown below.

$$\text{STP (Sum of all TP)} = (110+180+215+175) = 680;$$

$$\text{SFP (Sum of FP)} = (5+5+7+9) = 26$$

$$\text{SFN (Sum of FN)} = (0+8+8+0) = 16;$$

$$\text{Global Precision} = \text{STP} / (\text{STP} + \text{SFP}) = 0.9631$$

$$\text{Global Recall} = \text{STP} / (\text{STP} + \text{SFN}) = 0.977;$$

$$\text{Micro-F1} = 0.97;$$

$$\text{Average Precision} = (\text{sum of all precisions}) / 4 = 0.9622;$$

$$\text{Average Recall} = (\text{sum of all Precisions}) / 4 = 0.9803; \text{Macro-F1} = 0.971.$$

For comparison purposes, the above metrics are also calculated for other deep learning models (RF, SVM, and KNN) similarly to those calculated for XGBoost as shown in Table3. XGBoost classifier demonstrates good performance to detect smart contract vulnerabilities compared with RF, SVM, and KNN as shown in Figure 9.

The novelty of the proposed paper can be observed in Table 4, as it combines the two important features, which are contract snippets from [7,10] and the n-gram feature from [5], then an XGBoost ensemble model inspired by [5,20] to improve detection accuracy results and also work done towards auto SC generation inspired by [21,22].

Table 2. F1 Score calculation for XGBoost from confusion matrix elements.

XG Boost	TP	FP	FN	TN	Precision	Recall	F1-Score
C1	110	5	0	205	0.9565	1	0.9777
C2	180	5	8	137	0.9729	0.9574	0.9651
C3	215	7	8	120	0.9684	0.9641	0.9662
C4	175	9	0	196	0.9510	1	0.9749

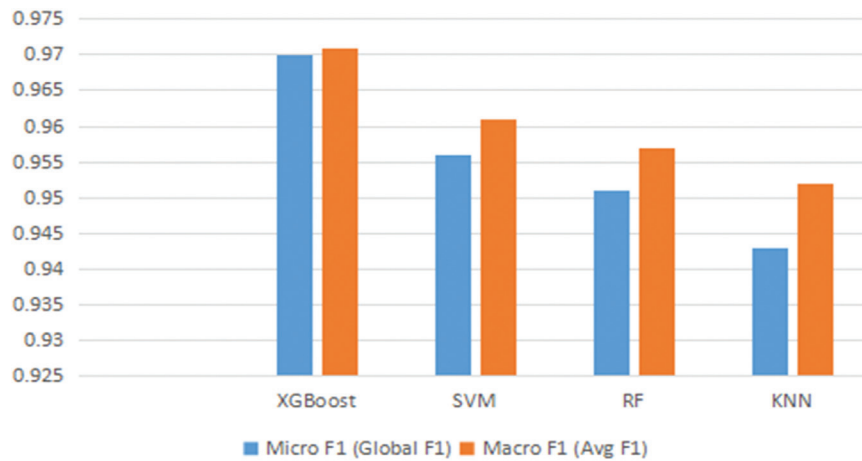


Fig. 9. Performance Comparison graph

Table 3. Micro-F1 and Macro-F1 calculation for Machine Learning Models.

ML MODEL	FI-SCORE				Micro_F1 (Global F1)	Macro_F1 (Avg F1)
	DOS (C1)	Unchecked Exception (C2)	Reentrancy (C3)	Tx_Origin (C4)		
XGBoost	0.977	0.965	0.966	0.974	0.97	0.971
RF	0.962	0.965	0.944	0.921	0.951	0.957
SVM	0.951	0.947	0.962	0.954	0.956	0.961
KNN	0.942	0.956	0.937	0.968	0.943	0.952

Table 4. Comparing proposed model with Existing Papers

	Vulnerabilities Detection	Features Used	ML Model Used	Work Done Towards Auto SC Generation
Liu Zhenguang et al [10]	Reentrancy	Contact Snippets	Bi-LSTM	No
WeiWang et al [5]	Reentrancy, Timestamp, Overflow, Underflow, Callstack, TOD	Bi-gram	XGBoost	No
Peng Qian et al [7]	Reentrancy, Timestamp, Infinite Loop	Pattern (Snippets) Extraction, Graph construction	CNN	No
Zhang L [20]	Reentrancy, Timestamp, Overflow, Underflow, Callstack, TOD	Information Graph	Ensemble Learning	No
C. K. Frantz et al[21]	--NA--	ADICO-Solidity DSL	--NA--	YES
Luciano B et al [22]	--NA--	Petri net graphs	--NA--	YES
Proposed Model	Re-entrancy, DOS, Unchecked external call, Origin of Transaction	Contact Snippets, N-gram, BPMN-SOL Compiler	XGBoost	YES

5. CONCLUSION & FUTURE WORK

This paper proposed work towards auto-smart contract generation and smart contract vulnerability detection models to identify specific security-related vulnerabilities using the XGBoost multi-label classification model. As shown in Figure 9, the proposed XGBoost model produced a 2% better average F1 score (2% better results than RF and KNN, and 1% better results than SVM), Compared with RF, SVM, and KNN deep learning models, by combining the best two features (called contract snippets and n-grams) from the literature to prepare a data set for the XGBoost model to detect SC

vulnerabilities, which are Denial of Service (DOS), Unchecked external call, Re-entrancy, and Origin of Transaction. This paper also makes use of BPMN and BPMN-SOL tools to initiate the work towards auto-smart contract generation. The limitation of this work is that the BPMN-SOL tool produces the smart contract template, where developers have to put effort into preparing the final smart contract. In our future work, we will concentrate on improving SC template quality, again trying to make it a fully automated conversion, and also try to improve SC vulnerability detection accuracy results using advanced deep learning concepts.

6. REFERENCES

- [1] W. Wang, D. T. Hoang, P. Hu, Z. Xiong, D. Niyato, "A survey on consensus mechanisms and mining strategy management in Blockchain networks", *IEEE Access*, Vol. 7, 2019, pp. 22328-22370.
- [2] A. Bruyn, Shanti, "Blockchain an introduction", University Amsterdam, 2017, pp. 1-43.
- [3] A. P. Joshi, M. Han, Y. Wang, "A survey on security and privacy issues of Blockchain technology", *Mathematical Foundations of Computing*, Vol. 1, No. 2, 2018, pp. 121-147.
- [4] Cointelegraph, <https://cointelegraph.com/bitcoin-for-beginners/how-Blockchain-technology-works-guide-for-beginners#where-can-Blockchain-be-used> (accessed: 2022)
- [5] Wei. Wang, Song. Jingjing , Xu. Guangquan, Li. Yidong, Hao. Wan, Su. Chunhua, "ContractWard: Automated Vulnerability Detection Models for Ethereum Smart Contracts", *IEEE Transactions on Network Science and Engineering*, Vol. 8, No. 2, 2021, pp. 1133-1144.
- [6] Liao. Jian-Wei, Tsai. Tsung-Ta, "SoliAudit: Smart Contract Vulnerability Assessment Based on Machine Learning and Fuzz Testin", *Proceedings of the Sixth International Conference on Internet of Things: Systems, Management and Security*, 2019, pp. 458-465.
- [7] P. Qian, W. Xun, "Combining Graph Neural Networks with Expert Knowledge for Smart Contract Vulnerability Detection", *IEEE Transactions on Knowledge and Data Engineering*, 2021. (in press)
- [8] S. S. Gupta, O. Yew-Soon, "Learning Approach to Detecting Security Threats", *Proceedings of ACM*, New York, NY, USA, 2019.
- [9] P. Momeni, Y. Wang, R. Samavi, "Machine Learning Model for Smart Contracts Security Analysis", *Proceedings of the 17th International Conference on Privacy, Security and Trust*, Fredericton, NB, Canada, 26-28 August 2019.
- [10] L. Zhenguang, H. Qingming, "Towards Automated Reentrancy Detection for Smart Contracts Based on Sequential Models." *IEEE Access*, Vol. 8, 2020, pp. 19685-19695.
- [11] W. J.-W. Tann, X. J. Han, S. S. Gupta, O. Yew-Soon, "Towards Safer Smart Contracts: A Sequence Learning Approach to Detecting Security Threats", *Proceedings of ACM*, New York, NY, USA, 2019.
- [12] B. Mueller, "ConsenSys/Mythril", [http:// github.com/ ConsenSys/mythril](http://github.com/ConsenSys/mythril) (accessed: 2020)
- [13] O. López-Pintado, B. García-Bañuelos, M. Dumas, I. Weber, A. Ponomarev, "Caterpillar: A Business Process Execution Engine on the Ethereum Blockchain", *Software: Practice and Experience*, 2018, pp:1-45.
- [14] M. Feng, Wang. Zhuoyi, "VSCL: Automating Vulnerability Detection in Smart Contracts with Deep Learning", *Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency*, Sydney, Australia, 3-6 May 2021.
- [15] K. L. Narayana, K. Sathiyamurthy, "Automation and smart materials in detecting smart contracts vulnerabilities in Blockchain using deep learning", *Materials Today: Proceedings*, 2021. (in press)
- [16] G. Chhaya, N. S. Gill, P. Gulia, "SSDT: Distance Tracking Model Based on Deep Learning", *International Journal of Electrical and Computer Engineering Systems*, Vol. 13, No. 5, 2022.
- [17] M. R. Isa, M. A. Khairuddin, "SIEM Network Behaviour Monitoring Framework using Deep Learning Approach for Campus Network Infrastructur", *International Journal of Electrical and Computer Engineering Systems*, 2021.
- [18] S. A. Baharudin, A. Lajis, "Deep Learning Approach for cognitive competency assessment in Computer Programming subject", *International Journal of Electrical and Computer Engineering Systems*, 2021.
- [19] L. Yiping, X. Jie, C. Baojiang, "Smart Contract Vulnerability Detection Based on Symbolic Execution Technology", *Proceedings of the China Cyber Security Annual Conference*, 2021, pp. 193-207.
- [20] L. Zhang, J. Wang, W. Wang, "A Novel Smart Contract Vulnerability Detection Method Based on Information Graph and Ensemble Learning", *Sensors*, Vol. 22, No. 9, 2022.
- [21] C. K. Frantz, M. Nowostawski, "From Institutions to Code: Towards Automated Generation of Smart

- Contracts”, Proceedings of the Workshop on Engineering Collective Adaptive Systems, co-located with SASO, Augsburg, 2016.
- [22] B. Luciano, A. Ponomarev, “Optimized Execution of Business Processes on Blockchain”, arXiv:1612.03152v1, 2016.
- [23] BPMN-SOL Compiler, <https://github.com/signavio/BPMN-Sol> (accessed: 2022)
- [24] BPMN-SOL Compiler, <https://github.com/shaunazzopardi/bpmn-to-solidity> (accessed: 2022)
- [25] BPMN, <https://www.visual-paradigm.com/guide/bpmn/what-is-bpmn/> (accessed: 2022)
- [26] BPMN, <https://www.lucidchart.com/pages/bpmn>. (accessed: 2022)
- [27] Smart Contract Dataset, <https://swcregistry.io/docs/SWC-107#modifier-reentrancy-fixedsol> (accessed: 2022)
- [28] Smart Contract Dataset, <https://github.com/smartbugs/smartbugs> (accessed: 2022)
- [29] Classification-loss-metrics, <https://peltarion.com/knowledge-center/documentation/evaluation-view/classification-loss-metrics/f1-score> (accessed: 2022)
- [30] Ethereum Smart Contract Best Practices, https://consensys.github.io/smart-contract-best-practices/known_attacks/ (accessed: 2022)

An empirical study on English-Mizo Statistical Machine Translation with Bible Corpus

Original Scientific Paper

Chanambam Sveta Devi

Department of Computer Science,
Assam University, Silchar, Assam, India
chsveta91@gmail.com

Loitongbam Sanayai Meetei

Department of Computer Science,
National Institute of Technology, Silchar, Assam, India
loisanayai@gmail.com

Bipul Syam Purkayastha

Department of Computer Science,
Assam University, Silchar, Assam, India
bipul_sh@hotmail.com

Abstract – Machine Translation (MT) is the process of automatically converting the text or speech in one natural language to another language with the help of a machine. This work presents a Bidirectional Statistical Machine Translation (SMT) system of an extremely low resource language pair Mizo-English, built in a low resource setting. A total of 30800 sentences are collected from the English Bible dataset and manually translated to Mizo by a native linguistic expert to generate the English-Mizo parallel dataset. After subjecting to various pre-processing steps, the parallel dataset is used to build our MT system using MOSES tools. Our framework uses different tools, such as GIZA++ for creating the Translation Model (TM) and IRSTLM to determine the probability of the target model. The quality of our MT system is evaluated using two automatic evaluation metrics: BLEU and METEOR. Our MT systems are also manually evaluated using two parameters: adequacy and fluency.

Keywords: Low resource, Statistical Machine Translation, Language Model, Translation Model, English, Mizo, Moses

1. INTRODUCTION

Machine translation (MT) is becoming a driving factor for every sector, such as academia and industry, as the demand for global communication increases. MT is an application of Natural Language Processing (NLP), and its development is correlated to data availability. Corpus-Based MT systems utilize a self-educated way of source-to-target study mapping. Corpus-Based MT systems include Example-Based Machine Translation (EBMT) [1,2], Statistical Machine Translation (SMT) [3,4] and Neural Machine Translation (NMT) [5,6]. SMT system's performance is correlated to the amount of parallel text (a text and its translation into another language) used to train the system. The basic principle of SMT is to employ huge parallel text in the training model to produce a better translation. The inability to correctly utilize information, computational complexity, and the need for many separate independently trained parts are only a few of the pri-

mary limitations of SMT systems. Although employing language-independent intermediate representation for translation is intriguing (as in the instance of Interlingua-based Machine Translation), linguistic diversity poses a challenge and rules out its viability. Most people now use the Internet for personal and professional activities, and the distinction between the real and digital worlds is becoming increasingly blurred. This digital world is becoming a reality for most of the world's population, and information security is becoming as important as physical security.

Mizo is the lingua franca of Mizoram, a northeastern state of India, and is spoken by around one million people. Mizo is the dominant language spoken by the resident people of Mizoram. Mizo is a Kuki-chin language, a branch of the Sino-Tibetan language, and belongs to the Tibeto-Burman family. The word order in English and Mizo is different; English follows SVO (Subject-Verb-

Object), and Mizo follows OSV (Object-Subject-Verb); however, Mizo sometimes follows SVO like English [7]. Furthermore, the second person pronoun "you" is used in English to represent both the singular and plural, whereas the Mizo language has unique phrases for this (for singular "I" and plural "in"). Although English rigorously maintains the order in which words must appear to construct a meaningful sentence, Mizo does not. English and Mizo are hardly related; however, both languages use the same Roman script. Prefixes and suffixes are affixes related to language morphology. Some of the Mizo language prefixes include "in," "ti," and "inti," and depending on the sentence, "ti" is occasionally adjusted and used as "tih" [7]. When suffixes are added after the stem word in Mizo, they can affect the part of speech, similar to how suffixes in English can change a verb into a noun. Furthermore, Mizo is a tonal language, which means that a word with various tones might have distinct meanings. There are eight tones in the language, four of which are reduced and four of which are long.

In this work, a manually translated parallel dataset of English to Mizo is built with the help of a native linguistic expert. The dataset is then used to train Statistical Machine Translation (SMT) systems with various settings. The MT systems are evaluated using automatic evaluation metrics: BLEU (Bilingual Evaluation Understudy) [8], METEOR (Metric for Evaluation of Translation with Explicit ORdering) [9], and F-measure. System-generated translations were subjected to both human and automated examinations to assess the efficacy of statistical techniques in the context of the Mizo language. The significant findings of this work are:

- Building an English-Mizo Bible corpus,
- Evaluate the system performance with the n-gram phrase-base language model
- Automatic evaluation of SMT in terms of BLEU and METEOR.
- Manual evaluation in terms of adequacy and fluency with the help of a native linguistic expert

We organized the paper in the following way: Section 2 discusses the previous works on the MT Problem. Section 3 describes the overview of SMT architecture. Then, Section 4 illustrates the details of our corpus and preprocessing step. The experimental findings and evaluation of our system are discussed in Section 5, followed by the conclusion and future works in Section 6.

2. RELATED WORKS

Machine Translation is the early application of NLP, which started its journey in 1959 but was performed in 1980 in India. Some of the MT systems in the Indian Language include Sampark[10], Mantra[11], and AnuBharti[11]. Mantra is an MT system designed for the Rajya Sabha Secretariat at C-DAC, Bangalore. An MT system called "AnglaBharti" was built using the rule method and the generalized form of the lexicon. It was created

in the year 1991 at IIT Kanpur. Following "AnglaBharti," another MT system called "AnuBharti" was built by the same organization in 2004 [11]. This system has been used to translate Hindi to English. IndicTrans is recent work on the MT system for Indian languages trained on the Samanantar dataset [12]. Furthermore, there has been reports of MT for other Indian languages focusing on low resource scenario such as Khasi [13], Hindi [14-15] and Manipuri [16-18].

Previous Natural Language Processing (NLP) study on the Mizo language includes an analysis on post-editing effort required to build English-Mizo parallel dataset [19], a Multi-Word Expression (MWE) for Mizo language [20], identifying criteria for recognition of Name Entity Classes in Mizo language [21], resource building and POS tagging for the Mizo language [22]. The preliminary study of POS tagging in the Mizo language [23] addressed the distinctive characteristics of the Mizo language and the limitations of the Mizo tagging system. The framework of MT systems for English to Mizo needs more work. [24] discussed the development of various applications of NLP for Mizo Language and the pre-processing steps for English to Mizo SMT system. [25] trained an NMT system for English to Mizo on a parallel corpus of 10,675 sentences and evaluated it on a test dataset of 100 sentences. The author reported that the MT system prediction is reasonable based on fluency but worse on accuracy. [7] conducted a study to evaluate the English to Mizo NMT system on several test datasets from different domains. [26] conducted a study on English to Mizo MT systems (SMT and NMT) on a training dataset collected from various online sources. [27] extended the work of [26] with additional training dataset of 31,764 parallel sentences and evaluated their systems on three tests dataset of sizes 100, 100, and 798 sentences. The author trained their systems with PB-SMT and NMT (LSTM, BiLSTM, and Transformer). The NMT-Transformer model was reported to outperform their baseline system.

3. STATISTICAL MACHINE TRANSLATION MODEL

This section discusses our approach for English to Mizo translation. Statistical Machine Translation is a machine translation system that uses a corpus based on the Noisy Channel concept. We use MOSES [28] to implement our English to Mizo MT system on a Bible parallel corpus. The statistical technique may be categorized as Empirical or Corpus-based machine translation, which necessitates a sizeable parallel text corpus to produce a high-quality translation. The SMT technique gives a solution to ambiguity concerns in natural languages. Some advantages of the Statistical model are that it is simple to create and run, requires little language skills to extract knowledge from a corpus, lowers human effort, and saves time [28]. SMT aims to produce the target sentence from the source sentence using the parallel corpus. Fig. 1 shows the outline of the Statistical Machine Translation system.

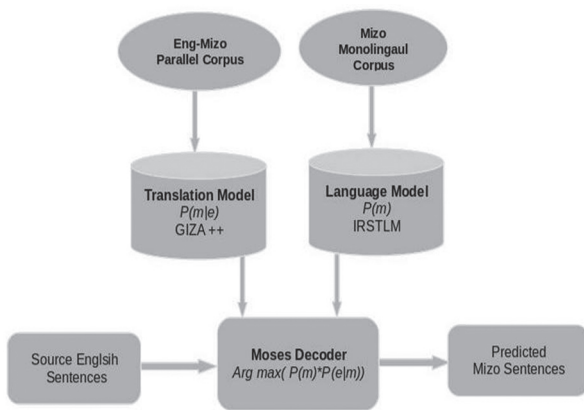


Fig. 1. Architecture of Statistical Machine Translation

The SMT architecture consists of three parts, namely, Language Model, Translation Model and Decoder.

A language model computes the probability of a sentence using an n-gram model. A language model may be thought of as a computation of the probability of a single word given all of the words that come before it in a sentence. It is divided into the conditional probability product. Using the chain rule, the probability of a phrase t , $P(t)$ is divided into the probability of individual words $\{w_1, w_2, w_3, \dots, w_n\}$, $P(w)$ as follows:

$$\begin{aligned}
 P(t) &= P(w_1, w_2, w_3, \dots, w_n) \\
 &= P(w_1)P(w_2|w_1)P(w_3|w_1w_2) \\
 &P(w_4|w_1w_2w_3)\dots P(w_n|w_1w_2 \dots w_{n-1})
 \end{aligned}$$

The Translation Model aids in calculating the conditional probability $P(m|e)$. It is the probability assigned to any pair of target sentence e and source sentence m . The parallel corpus of target-source pairings is used to train it. The process of computing the translation model is divided into smaller units, such as words or phrases, and their probabilities are learned. The translation of the source sentence is assumed to be generated word by word from the source. The translation of a target statement is as follows:

*(Ram is Riding his Bicycle /
Ram chuan a thirsakawr a khalh)*

The sentence pair having the possible alignment is given as,

*(Ram is Riding his Bicycle /
ram (1) chuan a (2) thirsakawr (5) a khalh (3,4).*

A variety of alignments are conceivable. To keep things simple, the translation model is aligned word by word. Consider the set of alignment by $B(e, m)$. If the length of the target m is l and the length of the source e is n , then there are $l \times n$ different connection of all possible alignment for each target position are equally likely, so the order of words in m and e has no effect on $P(m|e)$ and the likelihood of $(m|e)$ can be expressed as conditional probability $P(m, a|e)$ as $P(e|m) = \sum (e, a|m)$. The total is more than the element of the alignment set $B(e, m)$.

Decoder: To find the best translation from the given source sentence in target language by statistical model that compute the probability of language model and the probability of translation model. The $P(e, m)$ is the total possible outcome of the probability of alignment e and m . Now, we need to search for a pair (m, a) which $P(m, a|e)$ is maximize. By Bayes theorem, finding (m, a) which maximize $P(m, a|e) = P(e, a|m) * P(m)$. The early phrase-based statistical decoder model use the greedy hill-climbing algorithm [36], whereas the Moses decoder of phrase-base statistical model uses a beam search algorithm [37].

4. BUILDING CORPUS AND PRE-PROCESSING

This section discusses our corpus collection and data preparation.

For the experiment, 30,800 sentences in English are collected from the Bible monolingual corpus [29]. The collected sentences are manually translated to Mizo by the native speakers. Following is a sample example of the translation of an English sentence to Mizo:

English: *for god shall cast upon him, and not spare: he would fain flee out of his hand.*

Mizo: *zahngai lovin a nuai ang a, a thiltihtheihna hmaah chuan a kat rawk rawk ang.*

In the pre-processing step, non-ASCII special characters are removed from the parallel corpus to remove noise. The cleaned corpus is then tokenized with Moses Tokenizer [30]. Table 1 show the statistics of our experimental dataset.

Table 1. Statistics of English-Mizo parallel dataset

Language	Number of types	Number of unique types
English	920948	21235
Mizo	904484	13360

5. EXPERIMENTAL SETUP AND RESULTS

This section describes our experimental design and the evaluation of our English to Mizo MT systems. Table 2 contains essential information regarding the data utilized for the MT system.

Table 2. Data split up of the experimental dataset

Types	Number of sentences
Training	28300
Tuning	1500
Testing	1000

The MT systems are trained using the Moses toolkit. GIZA++ Toolkit [31] is used to generate the word alignment of the parallel corpora in both directions. IRSTLM [32] toolkit was used to train the language models (word and phrase-based). Tuning is performed by decoding and minimum error rate training (MERT) [33]. The alignments are then integrated using the grow-

diag-final and heuristic to produce a symmetric word alignment model [35].

5.1. AUTOMATIC EVALUATION

We examine the performance of our MT systems in terms of BLEU [8] and METEOR [9]. BLEU is an n-gram precision parameter, with higher values indicating better performance. METEOR rewards recollection by altering the BLEU brevity penalty, considers higher order n-grams to reward word order matches, and use arithmetic rather than geometric averaging.

We separately trained the MT system with three language models for the basic system (5, 4, and 3-grams standard phrase-based language models).

Table 3. SMT-LM System (English-Mizo)

n-gram	BLEU	METEOR	F-measure
3-gram	16.99	0.20	0.45
4-gram	17.36	0.21	0.46
5-gram	18.71	0.21	0.46

Table 4. SMT-LM System (Mizo-English)

n-gram	BLEU	METEOR	F-measure
3-gram	18.04	0.23	0.50
4-gram	19.25	0.23	0.51
5-gram	19.44	0.24	0.51

Table 3 and Table 4 shows the performance of English to Mizo and Mizo to English SMT systems in terms of BLEU, METEOR and F-measure. The result shows that the MT system with a 5-gram order of LM outperforms the 3-gram and 4-gram for both the English to Mizo and Mizo to English directions. The highest BLEU score for the English to Mizo SMT system is 18.71, and for the Mizo to English SMT system is 19.44.

Our results show similar performance of the MT system in terms of METEOR. For English to Mizo, the SMT system trained with the 4-gram and 5-gram order of LM achieve the same score of 0.21 and for Mizo to English, the SMT system trained with the 5-gram order of LM achieve the highest score of 0.24.

5.2. ANALYSIS OF THE MT SYSTEM BASED ON SENTENCE LENGTH

We also conducted an analysis of SMT systems based on the length of the sentences. Fig. 2 and Fig. 3 show the results from our English to Mizo and Mizo to English SMT systems, respectively. 3o-LM, 4o-LM, and 5o-LM are the SMT systems trained with 3-gram, 4-gram, and 5-gram orders of the Language Model, respectively.

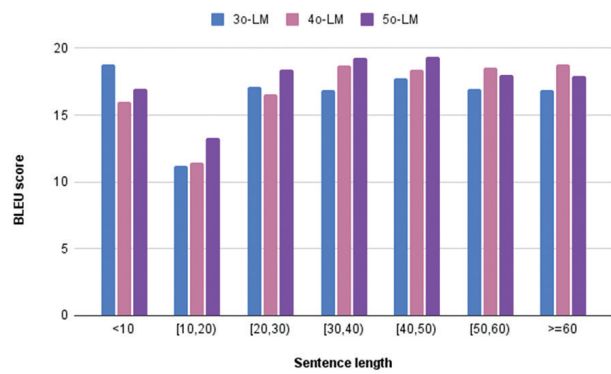


Fig. 2. Evaluation of our English to Mizo SMT systems

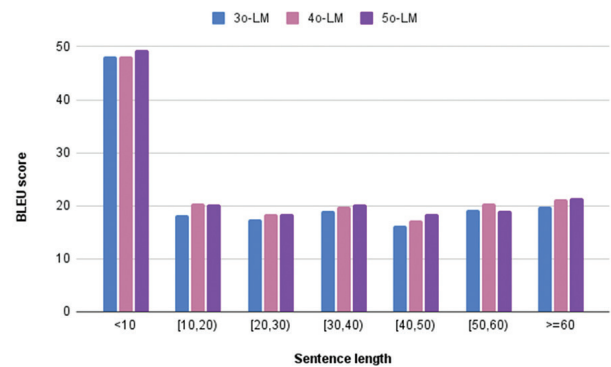


Fig. 3. Evaluation of our Mizo to English SMT systems

English to Mizo: The results from Fig.2 show that for short sentences with a length of less than 10, the 3o-LM MT system significantly outperforms the other MT systems. In the case of sentences of length greater than and equal to 50, the 4o-LM MT system performs best. However, overall the 5o-LM MT system is observed to perform better than the 3o-LM and 4o-LM SMT systems.

Mizo to English: The results from Fig.3 show that the 5o-LM MT system significantly outperforms the other MT systems in most cases.

From the above experimental results, we observe that unlike the results of English to Mizo SMT systems, the performance of the Mizo to English SMT system differs significantly for the sentences with lengths less than ten and the sentences with lengths greater than and equal to 10. The Mizo to English SMT is observed to be more robust for short sentences compared to English to Mizo SMT systems.

5.3. MANUAL EVALUATION

Manual evaluation is the best way of judging the quality of MT systems. Linguistic experts judge the output of MT quality based on the two-parameter: Adequacy and Fluency. Adequacy measures the amount of translation meaning of reference translation, which is included in a candidate translation.

Fluency is considered as well-formed grammatical sentences of the target language [35]. The scale used to measure the Adequacy and Fluency of our MT systems is shown in Table 5.

Table 5. Scale for Adequacy and Fluency

Scale	Adequacy	Fluency
5	All meaning	Flawless language
4	Most meaning	Good language
3	Much meaning	Non-native language
2	Little meaning	Disfluent language
1	None	Incomprehensible

Table 6. SMT-LM System (Mizo-English).

Order of N-gram LM	Adequacy	Fluency
3-gram	2.4	2.1
4-gram	3.3	3.1
5-gram	3.6	3.3

Table 7. SMT-LM System (English-Mizo)

Order of N-gram LM	Adequacy	Fluency
3-gram	3.1	2.0
4-gram	3.6	3.3
5-gram	3.9	3.6

Table 6 and Table 7 shows the Adequacy and Fluency score of our MT systems evaluated by the native linguistic experts. The results obtained from the manual evaluation complemented our findings from the automatic evaluation.

Following are the sample outputs from our SMT systems:

English to Mizo sample input-output:

English: all the cities of the children of aaron , the priests , were thirteen cities with their suburbs .

Reference: arona thlah , puithiamho khawpui zawng zawng chu khawpui sâwm leh pathum a ni , a daivêlte nêh .

3o-LM: khawpui zawng zawng chu arona thlah , puithiam chu khaw sâwm leh a daivêlte nêh .

4o-LM: khawpui zawng zawng chu arona thlah puithiamte chu an ni , " a ti a , khawpui sâwm leh pathum a ni , a daivêlte nêh .

5o-LM: khawpui zawng zawng chu arona thlah , puithiam chu khawpui sâwm leh pathum a ni , a daivêlte nêh .

Mizo to English sample input-output:

Mizo: an vaiin an ei a , an tlai ta hlawm a , an sem bâng chu bawm sarifah an dah khat a .

Reference: and they did all eat , and were filled : and

they took up of the broken meat that was left seven baskets full .

3o-LM: and they were they did eat , and were filled , and they took up the ark was left seven baskets .

4o-LM: and they were they did eat , and were filled : and they took up of the broken meat that was left seven baskets .

5o-LM: and they were they did eat , and were filled : and they took up of the broken meat that was left seven baskets .

6. CONCLUSION

This paper discusses corpus standardization and a strategy for developing standardized datasets. The parameter for generating the best performance for bi-directional English-Mizo SMT is also determined. An analysis of the n-gram order of the language model for the SMT system is carried out in this work. The advantage of this method is that it employs the exact phrases found in the translation table and those included in the target part of each entry. The extensive experiments on English-to-Mizo and Mizo-to-English translation indicate that the phrase-based language model can increase the quality of the SMT system. The systems are assessed using the automated scoring methodologies BLEU and METEOR score and manual evaluation by linguistic experts. The SMT systems trained with 5-gram order of Language Model outperform the other MT systems by achieving a BLEU score of 18.71 for English to Mizo and 19.44 for Mizo to English. The automatic evaluation results show that the performance of the MT system increases with the n-gram order of LM. In the future, we will analyze the result of the English-Mizo MT systems by increasing the corpus size from different domains.

7. REFERENCES:

- [1] E. Sumita, H. Iida, "Experiments and prospects of example-based machine translation", Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, June 1991, pp. 185-192.
- [2] M. R. Costa-Jussà, J. Centelles, "Description of the Chinese-to-Spanish rule-based machine translation system developed using a hybrid combination of human annotation and statistical techniques", ACM Transactions on Asian and Low-Resource Language Information Processing, Vol. 15, No. 1, 2015, pp. 1-3.
- [3] P. Koehn, F. J. Och, D. Marcu, "Statistical phrase-based translation", Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Vol. 1, 2003, pp. 48-54.

- [4] T. D. Singh, S. Bandyopadhyay, "Bidirectional Statistical Machine Translation of Manipuri English Language Pair using Morpho-Syntactic and Dependency Relations", *International Journal of Translation*, Vol. 23, No. 1, 2011, pp. 115-37.
- [5] D. Bahdanau, K. Cho, Y. Bengio, "Neural machine translation by jointly learning to align and translate", arXiv:1409.0473, 2014.
- [6] L. Rahul, L.S. Meetei, H.S. Jayanna, "Statistical and Neural Machine Translation for Manipuri-English on Intelligence Domain", *Advances in Computing and Network Communications*, 2021, pp. 249-257.
- [7] Z. Thihlum, V. Khenglawt, S. Debnath, "Machine Translation of English Language to Mizo Language", *Proceedings of the IEEE International Conference on Cloud Computing in Emerging Markets*, 2020, pp. 92-97.
- [8] K. Papineni, S., Roukos, T. Ward, W.J. Zhu, "Bleu: a method for automatic evaluation of machine translation", *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, July 2002, pp. 311-318.
- [9] M. Denkowski, A. Lavie, "Meteor universal: Language specific translation evaluation for any target language", *Proceedings of the 9th workshop on statistical machine translation*, June 2014, pp. 376-380.
- [10] S. K. Dwivedi, P. P. Sukhadeve, "Machine translation system in indian perspectives", *Journal of computer science*, Vol. 6, No. 10, 2010, pp. 1111-1116.
- [11] S. K. Naskar, S. Bandyopadhyay, "Use of machine translation in India: Current status", *Proceedings of Machine Translation Summit X: Posters*, 2005, pp. 465-470.
- [12] G. Ramesh et al. "Samanantar: The largest publicly available parallel corpora collection for 11 indic languages", *Transactions of the Association for Computational Linguistics*, Vol. 10, 2022, pp.145-162.
- [13] T. D. Singh, A. V. Hujon, "Low resource and domain specific english to khasi smt and nmt systems", *Proceedings of the International Conference on Computational Performance Evaluation*, 2020, pp. 733-737.
- [14] L. S. Meetei, T. D. Singh, S. Bandyopadhyay, "WAT2019: English-Hindi translation on Hindi visual genome dataset", *Proceedings of the 6th Workshop on Asian Translation*, 2019, pp. 181-188.
- [15] L. S. Meetei, T.D. Singh, S. Bandyopadhyay, "Low Resource Multimodal Neural Machine Translation of English-Hindi in News Domain", *Proceedings of the First Workshop on Multimodal Machine Translation for Low Resource Languages*, September 2021, pp. 20-29.
- [16] S. M. Singh, T. D. Singh, "An empirical study of low-resource neural machine translation of manipuri in multilingual settings", *Neural Computing and Applications*, 2022, pp. 1-22.
- [17] S. M. Singh, T. D. Singh, "Low resource machine translation of english-manipuri: A semi-supervised approach", *Expert Systems with Applications*, Vol. 209, 2022, pp. 118-187.
- [18] S. M. Singh, T. D. Singh, "Unsupervised neural machine translation for english and manipuri", *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, December 2020, pp. 69-78.
- [19] L. S. Meetei, T. D. Singh, S. Bandyopadhyay, M. Vela, J. van Genabith, "English to Manipuri and mizo post-editing effort and its impact on low resource machine translation", *Proceedings of the 17th International Conference on Natural Language Processing*, December 2020, pp. 50-59.
- [20] G. Majumder, P., Pakray, Z. Khiangte, A. Gelbukh, "Multiword expressions (MWE) for Mizo Language: literature survey", *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, April 2016, pp. 623-635.
- [21] J. Bentham, P. Pakray, G. Majumder, S. Lalbiaknia, A. Gelbukh, "Identification of rules for recognition of named entity classes in mizo language", *Proceedings of the 15th Mexican International Conference on Artificial Intelligence*, IEEE, Oct 2016, pp. 8-13
- [22] P. Pakray, A. Pal, G. Majumder, A. Gelbukh, "Resource building and parts-of-speech (pos) tagging for the mizo language", *Proceedings of the 14th Mexican International Conference on Artificial Intelligence*, October 2015, pp. 3-7.

- [23] M. V. Nunsanga, P. Pakray, M. Lalngaihtuaha, L. L. K. Singh, "Part-of-speech tagging in Mizo language: A preliminary study", *Data Intelligence and Cognitive Informatics*, Springer, 2021, pp. 625- 635.
- [24] C. S. Devi, B. S. Purkayastha, "Development of various applications of NLP for Mizo Language", *Recent Trends in Programming languages*, Vol. 7, No. 1, 2020, pp. 7-15.
- [25] C. S. Devi, B. S. Purkayastha, "Steps of Pre-processing for English to Mizo SMT System", *Proceedings of the International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, Singapore, July 2020, pp. 156-167.
- [26] A. Pathak, P. Pakray, J. Bentham, "English-Mizo machine translation using neural and statistical approaches", *Neural Computing and Applications*, Vol. 31, No. 11, 2019, pp. 7615-7631.
- [27] C. Lalrempuii, B. Soni, P. Pakray, "An Improved English-to-Mizo Neural Machine Translation", *Transactions on Asian and Low-Resource Language Information Processing*, Vol. 20, No. 4, 2021, pp. 1-21.
- [28] P. Koehn, "Moses, statistical machine translation system, user manual and code guide", 2010.
- [29] GNB:Bible you version homepage <https://www.bible.com/en-GB/bible/2163>.
- [30] P. Koehn et al. "Moses: Open source toolkit for statistical machine translation", *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions*, 2007, pp. 177-180.
- [31] F. J. Och, H. Ney, "A Systematic Comparison of Various Statistical Alignment Models," *Computational Linguistics*, vol. 29, No. 1, 2003, pp. 19-51.
- [32] M. Federico, N. Bertoldi, M. Cettolo, "IRSTLM: an open source toolkit for handling large scale language models", *Proceedings of the 9th Annual Conference of the International Speech Communication Association*, 2008.
- [33] F. J. Och, "Minimum error rate training in statistical machine translation", *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, July 2003, pp.160-167.
- [34] P. Koehn, F. J. Och, D. Marcu, "Statistical phrase-based translation", *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 2003, pp. 127-133.
- [35] M. S. Maučec, G. Donaj, "Machine Translation and the Evaluation of its Quality", *Recent Trends in Computational Intelligence*, Vol. 143, 2019.
- [36] O. Miles. "Statistical Machine Translation ", 2010, pp.912-915.
- [37] G. Ulrich, M. Jahr, K. Knight, D. Marcu, K. Yamada, "Fast and optimal decoding for Machine Translation", *Artificial Intelligence*, Vol. 154, No. 1-2, 2004, pp. 127-143.

Speaker Recognition Based on Mutated Monarch Butterfly Optimization Configured Artificial Neural Network

Original Scientific Paper

Dhana Lakshmi Namburi

Research Scholar, ANU, Guntur, Andhra Pradesh, India
Assistant Professor, CBIT, Hyderabad, Telangana, India
dhananm@gmail.com

Satya Sai Ram M

RVR & JC College of Engineering,
Chowdavaram, Andhra Pradesh, India
Msatyasairam1981@gmail.com

Abstract – Speaker recognition is the process of extracting speaker-specific details from voice waves to validate the features asserted by system users; in other words, it allows voice-controlled access to a range of services. The research initiates with extraction features from voice signals and employing those features in Artificial Neural Network (ANN) for speaker recognition. Increasing the number of hidden layers and their associated neurons reduces the training error and increases the computational process's complexity. It is essential to have an optimal number of hidden layers and their corresponding, but attaining those optimal configurations through a manual or trial and the process takes time and makes the process more complex. This urges incorporating optimization approaches for finding optimal hidden layers and their corresponding neurons. The technique involve in configuring the ANN is Mutated Monarch Butterfly Optimization (MMBO). The proposed MMBO employed for configuring the ANN achieves the sensitivity of 97.5% in a real-time database that is superior to contest techniques.

Keywords: Speaker recognition, Speaker verification, Speaker identification, Artificial Neural Network, Monarch Butterfly Optimization, Model configuration.

1. INTRODUCTION

Since a decade ago, academics and industry have paid increasing attention to speaker identification [1]. It is extensively used in applications, including security and surveillance, financial security, discriminative speaker embedding learning, voice authentication, forensic voice verification for suspect detection [2], electronic voice eavesdropping, voice conversion, and identity verification, as well as access control, biometrics authentication, mobile shopping, and mobile banking [3]. It essentially involves classifying unknown speakers based on their speech [4]. Speaker identification is the process of identifying a speaker sound based on a set of trained speaker sounds. In other words, speaker identification compares one user's voice profile with many other profiles and determines the best or exact match. Since speech signals are the primary means of communication, they constantly contain rich, relevant details, such as speakers' accents, gender, emotions, and other characteristics. As a result of these distinctive characteristics, researchers can distinguish

between speakers during phone calls, even when the speakers are not physically present [6] [8] [9].

Speaker Identification involves identifying unknown voices from a fixed set of known speakers. Therefore, it is called closed set identification. Based on the speech used for identifying the speaker, the systems can be grouped into text-dependent (fixed text is used for both training and testing phase) and text-independent (no fixed text). Out of the two types, text-independent speaker recognition is most challenging job. The error that can occur in speaker identification is false identification, which can be measured by sensitivity, which determines the correctness of the predictions. A high sensitivity model provides a more reliable result than a low sensitivity model in medical applications. Hence, the objective of this work is to build a model for text independent speaker recognition with improved recognition accuracy as well as sensitivity.

A variety of models, techniques, and algorithms are employed to identify speakers in recent literature, including Mel-frequency Cepstral Coefficients (MFCC)

and Linear Predictive Coding (LPC) [10], the Histogram Transform Model [11], and spatiotemporal sparse coding and hierarchical pooling [12]. In HT based SI systems achieves identification accuracy of 99.52% and is affected by H (random affine transformations). Increasing H improves the identification accuracy, but when H is higher than 400, the accuracy decreases instead. Similarly in Visual speaker identification and authentication by joint spatiotemporal sparse coding and hierarchical pooling archives higher identification accuracy with the increase of dictionary size K. As K increases results in very high computational complexity and large memory cost during the classifier training process.

Despite their effectiveness and accuracy, these traditional speaker identification methods have not been able to identify human voices effectively. A method based on Artificial Intelligence (AI) has been proposed by speech processing researchers to overcome this issue [13]. In recent years, AI technology has substantially enhanced both the recognition rate and robustness of speaker identification. As a result, the results produced by machine learning neural networks continue to support neural networks' use in speaker identification. Support vector machines (SVMs), artificial neural networks (ANNs), and K-nearest neighbours (KNNs) are among the methods used to identify speakers in literature. Among these, ANNs have proven effective in identifying speakers. Over the last three decades, ANNs have also been extensively studied and applied to classification, pattern recognition, regression, and forecasting.

Despite its numerous advantages, traditional ANNs still lack accuracy and performance. Therefore, placing the optimal number of hidden layers enhances traditional ANN performance. In recent years, MBO (Monarch Butterfly Optimization) procedures have been proposed in various literature [18] [19], so the research employs MBO. The search strategy of the basic MBO algorithm, on the other hand, readily slips into local optima, resulting in precocious convergence and low performance on many complicated optimization tasks. Scholars have made several enhancements to MBO in recent years to improve its effectiveness [20][21]. However, these techniques do have not a sufficient performance in view of convergence speed and accurate optimum solution. To solve the issues, this paper develops an Oppositional based strategy with the Cauchy distribution (Cd) technique in MBO is proposed. First, is the Opposition Based Learning model, which ensures the exploration of unique and opposing candidate solutions in the search space while the evolution process is ongoing in order to assess the better candidate solutions [22] [23]. Secondly, Cd as a mutation operator enriches the conventional performance MBO algorithm [24].

2. LITERATURE REVIEW

Daqrouq et al. (2015) [25] had proposed a speaker recognition system that utilizes a combination of for-

mants, wavelets Entropy, and neural network classifiers to identify vowels characteristics. The initial stage involved extracting five formants and seven Shannon entropy wavelet packets from the speakers' signals to build the speaker feature vector. In the next stage, these 12 feature extraction coefficients were utilized as inputs to feed-forward neural networks. The suggested technique performs well in speaker verification and identification tasks, according to the findings of the experiments. The results were shown to be superior to well-known classical speaker detection techniques.

Faragallah, Osama S et al. (2018) [26] had proposed MKMFCC–SVM is a robust noisy automated speaker identification (ASI) technique. It uses a support vector machine and the Multiple Kernel Weighted-MFCC (MKMFCC). In the face of noise or deterioration, experimental studies showed that the suggested MKMFCC–SVM ASI method gives a greater identification rate.

Chen et al. (2019) [27] had proposed a bi-level framework to mutually optimize session compensation and support vector machine (SVM) based classifier for speaker identification. Finally, the trials demonstrated that in the i-vector framework, the proposed techniques outperformed existing session compensation algorithms and classifiers.

de Abreu Campos et al. (2019) [28] had proposed an unsupervised learning technique such as RL-Sim and ReckNN for speaker retrieval and recognition. The method was organised around a framework that makes use of a rank-based formulation. The adoption of unsupervised learning algorithms over standard speaker identification approaches resulted in effectiveness enhancements of up to +56 percent on retrieval measures.

Safavi et al. (2018) [29] had proposed Automatic identification of the speaker, age group, and gender from children's speech. A number of classification techniques were examined, including the Gaussian Mixture Model–Universal Background Model, GMM–SVM, and i-vector established systems. As one might imagine, the mistake rate for speaker recognition lowers with age. However, the influence of age on gender and age-group documentation was more complicated, owing to the repercussions of adolescent. Finally, the ability of distinct bandwidths to identify speakers, age groups, and gender from children's speech was tested.

Devi et al., (2020) [30] had proposed a hybrid technique for Automatic Speaker Recognition that uses speech signals and an ANN to increase speaker prediction accuracy. The proposed ANN-based approach was designed based on Multilayer Perceptron (MLP) with Bayesian Regularization. In contrast to existing models, the suggested strategy was validated by performance assessment and classification accuracies. The authors claimed that the suggested method provided a nicer recognition rate and 93.33% accuracy was achieved.

Biswas et al. (2021) [31] had proposed a multi-layer

perceptron neural network to identify singers' voices. The trials for singer identifications were repeated five times in this study, and the analysis was carried out using feature extraction. Apart from the employment of the supervised learning approach with the insinuation of weight optimization, the effectiveness was found for the recognition of the novel and unidentified vocalist to be discovered. Finally, the study found that the identification was accurate to the tune of 99.29%.

Wang et al. (2019) [32] had proposed a MBO is a nature-inspired metaheuristic algorithm inspired by monarch butterfly migratory behaviour. As a result, the monarch butterfly's locations were updated in two ways. The offspring were first created (position updating) by the migration operator, which may be changed by the migration ratio. The butterfly regulating operator is then utilized to fine-tune the locations of other butterflies. In comparison to previous algorithms, the MBO technique convincingly demonstrated its capacity to discover increased function values on majority of the benchmark issues.

Chakraborty et al. (2019) [33] had proposed by using Oppositional Based Learning (OBL) and Dynamic Cauchy Mutation (DCM), an Enhanced Elephant Herding Optimization (EEHO) to address the multilevel image thresholding issue for image segmentation. OBL improves the performance of normal EHO by speeding up the convergence rate, whereas DCM prevents premature convergence. This proposed algorithm delivers capable performance equated to other methods.

3. PROPOSED METHODOLOGY

Speaker recognition is a method for recognizing who is speaking automatically by utilizing speaker-specific information included in voice waves. The voice signal contains critical information such as message content, language, speaker identification, emotion, personality, and so on. It permits voice-controlled access to various services.

To build a Speaker Identification (SI) System, the model parameters are regularly learned based on the features extracted from the speech samples in the training phase. Testing involves feeding the extracted features from unknown speech to the trained model to identify who is speaking. Most widely used feature extraction methods are MFCC and LPC. The MFCC is a popular feature in Automatic Speech Recognition (ASR) and is inferred following static (non-signal dependent) processing methods. A LPC gives a decent model of the speech signal. This is particularly valid for the quasi-steady state voiced regions of speech in which all-pole model of LPC give a good approximation to the vocal tract spectral envelope. Different classifiers are likewise accessible for SI namely Kernel Regression and K Nearest Neighbour (KNN), Support Vector Machine (SVM), Hidden Markov Model (HMM), Maximum Likelihood Classifier (MLC) and ANN.

This research initiates with extraction features from voice signals and employing those features to Artificial Neural Network (ANN) for speaker recognition. It is also possible to reduce the training error by increasing the number of hidden layers and their linked neurons, as well as increasing the complexity of the computational process. In ANN, the hidden layer is critical for identifying characteristics in the input data and using them to correlate between a given input and the proper output. A higher number of hidden layers increases the order of weights, and it helps to make a higher-order decision boundary. Similarly, increasing hidden layers would also increase the complexity of the model and sometimes lead to over-fitting. It is essential to have an optimal number of hidden layers and their corresponding, but attaining those optimal configurations through a manual or trial and process takes time and makes the process more complex. These urges incorporate optimization methods for recognizing optimal hidden layers and their corresponding neurons. The technique involves in configuring the ANN is MMBO; the process of integrating the opposition strategy and the Cd strategy to enhance the performance of traditional MBO. The study used 200 real-time speech signal datasets from 20 speakers, including eight female and twelve male voice signals for 10 words each. ANN operates with Levenberg–Marquardt (LM) as a training technique for speaker recognition. This research considers 80% of datasets for training and the remaining 20% for testing the configured model.

The flow diagram of the research work is shown in figure 1. It consists of two stages. Stage 1 is training the model; Stage 2 is testing the model. In training process, Features are extracted from training data set and these features are used to train the ANN. ANN architecture is modified by optimizing the number of hidden layers and number of hidden layer neurons by using different optimization techniques.

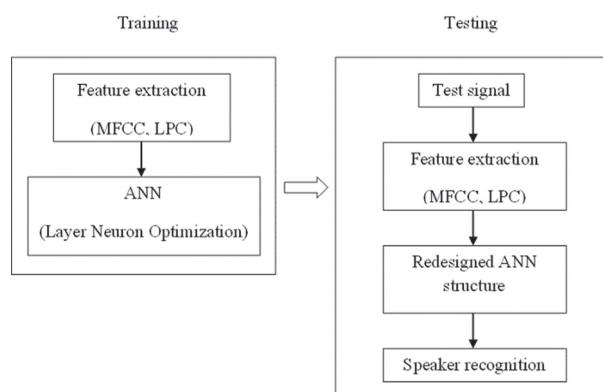


Fig. 1. Flow diagram of the process

3.1. FEATURE EXTRACTION

Feature extraction is important for speaker voice identification; this procedure is carried out using a couple of well-known algorithms named MFCC and LPC.

3.2. ARTIFICIAL NEURAL NETWORK (ANN)

ANN is adaptive and dynamic, learning and altering in response to each unique internal or exterior input. ANNs are employed in systems for sequence and pattern identification, data processing, robotics, and modeling.

3.2.1. MONARCH BUTTERFLY OPTIMIZATION (MBO)

Wang introduced the MBO algorithm in 2015, which is a type of swarm intelligence meta-heuristic procedure inspired by monarch butterfly migratory behavior. Individuals in MBO are updated through the migration and butterfly adjustment operations. When tackling global numerical optimization, the MBO outperforms numerous state-of-the-art optimization approaches. The migratory operator and the butterfly-adjusting operator are used to update the locations of monarch.

Initialization

The initialization of randomly generated solutions is the first step in every optimization approach. In the ranges of 1 to 5 and 1 to 30, the number of hidden layers and the values of their corresponding neurons are created at random. The length of the solution is determined by the value of the hidden layer created randomly in the first location. For example, if the value of the randomly allocated hidden layer is 3, the solution length will be 4 (3+1). Similarly, produce 10 solutions at random and feed them into the fitness process to assess the solution's strength. The primary purpose of opposition-based solution generation is to evaluate matching opposing estimates as a subsequent set of candidate solutions in order to improve the present candidate solution's approximation. An opposing candidate solution has been shown to have a higher likelihood of being nearer to the global optimal solution than a randomly picked candidate solution. Mathematically opposition based solution generation expressed as,

$$N_{(j,i)}^o = x_i + y_i - N_{(j,i)} \quad (1)$$

Let $N \in [a, b]$ be a real numbers; Where, N^o is the opposition based solution and N refers randomly generated solution and x_i, y_i refers the minimum and the maximum values respectively. The both randomly generated solution and the opposition based solution generations are fed in to fitness computation for process evaluation.

Fitness Function

The fitness approach to evaluate how well a solution performs in comparison to the overall amount of validation data.

$$\text{Fitness} = \frac{\text{Correctly Recognize}}{\text{Total Number of Validation Data}} \quad (2)$$

Migration Operator

The monarch butterfly migration between Lands 1 and 2 is expected to be updated by the migration op-

erator, with monarch butterflies solely belonging to subpopulations 1 and 2. Initially, $NP1 = \text{ceil}(p * NP)$ and $NP2 = NP - NP1$ may be used to compute the number of monarch butterflies in Lands 1 and 2.

Where NP represents the total number of monarch butterflies in Land 1, p denotes the monarch butterfly ratio in Land 1, and $\text{ceil}(y)$ represents the rounding of y to the nearest whole number larger than or equal to y. In this way, migration operator arranged as

$$y^{t+1}_{i,k} = \begin{cases} y^t_{r1,k} & | r \leq p \\ y^t_{r2,k} & | r > p \end{cases} \quad (3)$$

Wherever $y^{t+1}_{i,k}$ denotes the k^{th} element of y_i at generation t+1. Basically, $y^t_{r1,k}$ demonstrates the k^{th} element of y_{r1} at generation t, and $y^t_{r2,k}$ denotes the kth element of y_{r2} at generation t. t represents the current generation number. Monarch butterflies (r1 and r2) are arbitrarily selected from subpopulations 1 and 2. The condition variable (C_r) is found as follows:

$$C_r = \text{rand} * m_{tr} \quad (4)$$

The fundamental MBO technique is as follows: m_{tr} is the migration time frame, which is set to 1.2, and rand is a arbitrary number derivative from a uniform distribution.

Butterfly Adjusting Operator

The locations of monarch butterflies in subpopulation 2 are updated utilizing this operator. It may be updated as follows:

$$y^{t+1}_{j,k} = \begin{cases} y^t_{best,k} & | \text{rand} \leq p \\ y^t_{r3,k} & | \text{rand} > p \end{cases} \quad (5)$$

Where $y^{t+1}_{j,k}$ denotes the k^{th} element of y_j at generation t + 1; $y^t_{best,k}$ denotes the kth element of ybest at generation t, this indicates the finest monarch butterfly habitat in Lands 1 and 2. The $y^t_{r3,k}$ denotes the kth element of y_{r3} at generation t; the monarch butterfly r3 is randomly selected from subpopulation 2. If $\text{rand} > p$, there is a different development. If $\text{rand} > \text{BAR}$, the butterfly's position is also updated using Levy flying:

$$y^{t+1}_{i,k} = y^{t+1}_{j,k} + \alpha \times (dy - 0.5) \quad (6)$$

The variable BAR stands for the Butterfly Adjusting Rate; if BAR is less than a arbitrary value, the kth element of y_j at generation t+1 is changed, where is the weighting factor, as exposed in Equation (7).

$$\alpha = WS_{\max} / t^2 \quad (7)$$

WS_{\max} denotes the maximum walk step. In Equation (6), dy is the butterfly j walk step that Levy flight can consider.

$$dy = \text{Levy}(y^t_j) \quad (8)$$

Finally, the freshly formed butterfly with the best fitness is promoted to the next generation and replaced by its father; it is also eliminated to preserve population number.

Cauchy Distribution

The Cd is a separate updating technique that works in tandem with the migration and butterfly adjustment operators. This is the continuous probability distribution that has two parameters, x_0 and γ . x_0 is the location parameter, and γ is the scale parameter that defines the shape of the Cd. For instance, if a developed value is allotted to γ , the height of the peak of the Probability Density Function (PDF) will be smaller, and its width will be broader. On the other side, if a lesser value is allocated to γ , the height of the peak of the PDF will be higher, and its thickness will be narrower. The Cd's PDF may be described as follows.

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x - x_0}{\gamma} \right)^2 \right]} = \frac{1}{\pi} \left[\frac{\gamma}{(x - x_0)^2 + \gamma^2} \right] \quad (9)$$

The Cd's cumulative distribution function may also be described as follows.

$$F(x; x_0, \gamma) = \frac{1}{\pi} \arctan \left(\frac{x - x_0}{\gamma} \right) + \frac{1}{2} \quad (10)$$

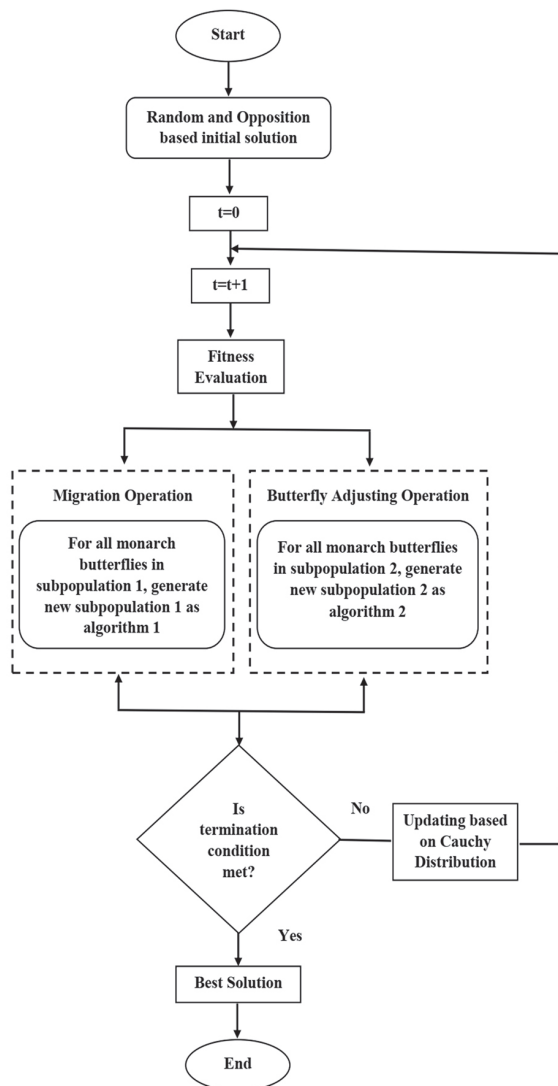


Fig. 2. Flow chart of Mutated Monarch Butterfly optimization

3.2.2 Comparison of Hyper parameters of all network structures

The default structure of ANN is comprised of one input layer, one hidden layer associated with ten neurons, and one output layer. Each neuron in the input layer is connected with the hidden layer neurons with random weight $w_{11}, w_{12} \dots w_{ij}$. Similarly, with the output layer. These initial random weights are adjusted based on the fed features. In Artificial Neural Networks, the Levenberg-Marquardt algorithm is most commonly used for training optimization, and the default transfer function is 'tansig'.

Table 1: Comparison of Hyper parameters of all network structures

Updating Best solution in every iteration	Updating parameters in Land 1 and Land 2	Initial Population (NP monarch butterflies)	Training Algorithm	Transfer function of the neurons	
No optimization techniques are employed for finding the optimum number of hidden layers and its associated neurons			trainlm	tansig	Default ANN
No change	Migration Operator and Butterfly Adjusting operator	Random solution generation	trainlm	tansig	MBO configured ANN
No change	Migration Operator and Butterfly Adjusting operator	Opposition-based solution generation	Trainlm	Tansig	OMBO configured ANN
The elements of best solution are further updated using Cauchy distribution	Migration Operator and Butterfly Adjusting operator	Random solution generation	trainlm	tansig	CMBO configured ANN
The elements of best solution are further updated using Cauchy distribution	Migration Operator and Butterfly Adjusting operator	Opposition-based solution generation	trainlm	tansig	MMBO configured ANN

In the basic MBO algorithm, local optima are easily reached, resulting in early convergence and poor performance. Using opposition-based learning (OBL) and the Cauchy distribution, this paper develops a novel MBO algorithm. Initially, OBL is used to create opposition-based populations from the original population. In opposition-based populations, the best individuals are selected and passed to the next generation, and

this process effectively prevents the MBO from falling into a local optimum.

In this context, the optimal number of hidden layers and their associated neurons can be determined. If N is the number of hidden layer neurons generated randomly, then No is the opposition-based solution expressed in equation 1.

Secondly, Cauchy distribution is introduced to improve the migration and butterfly adjustment operators. In every iteration, it helps to update the best solution to improve the convergence rate.

4. RESULTS

Migration operator and butterfly adjusting operator in the MBO algorithm ensure monarch butterflies' search directions. In addition, the migration operator and the butterfly adjusting operator can be executed simultaneously. One of the advantages of MBO algorithm is its simplicity and ease of implementation. However, MBO algorithm drawback is poor optimization efficiency in solving complex optimization problems, which can be seen in the following aspect. The monarch butterflies r1 and r2 are randomly selected from Subpopulation1 and Subpopulation2, respectively. A worse monarch butterfly may be selected to share its features with a better one, leading to the population degenerating. This can be overcome by using Opposition-based Learning method. If the OBL approach is introduced into the initialization of the MBO algorithm, it can produce the opposition-based population. Then, the better individuals are selected to participate in the evolution from the union of the original populations and the opposition-based populations. Further every element in the best solution after every iteration is also updated by Cauchy distribution. Thus, these two operations increase the population diversity and expands the exploration scope of MBO. Further, it contributes to faster rate of convergence and better accuracy as well as sensitivity.

Configuring ANN via MMBO techniques accomplished 97.5% sensitivity for speaker recognition. Opposition based solution generation parallel with random solution generation and Cd function elevates the sensitivity over Oppositional based MBO, Cd based MBO and traditional MBO. The suggested MMBO creates an ANN with three hidden layers, each of which has 19, 23, and 23 neurons. The investigation shows the performance of involved techniques through diverse measures. It is obvious from the graphs that proposed approach having better performance over other techniques. The table 1 exhibits the ANN model configuration from different techniques. All at once, the employed optimization techniques in configuring ANN model show three-hidden layers. Though, the hidden layers are same for employed techniques change in respective hidden neurons impact effectively on proposed approach. The entire execution procedure took place on the MATLAB R2015a.

Table 2. ANN model configuration from different techniques

Techniques	Input	Hidden Layers	Neurons	Neurons	Neurons	Output
MBO-ANN	61	3	20	22	23	1
OMBO-ANN	61	3	20	25	21	1
CMBO-ANN	61	3	30	20	23	1
MMBO-ANN	61	3	19	23	23	1

The performance of the strategies used when configuring ANN for speaker recognition is exposed in Fig. 2. The results show that MBO's use of ANN configuration to forecast speaker voice recognition is superior to contest strategies.

True Positive (TP) - Recognised person's voice correctly identified as recognised person

False Positive (FP) - Not-Recognised person's voice incorrectly identified as recognised person

True Negative (TN) - Not-Recognised person's voice correctly identified as not-recognised person

False Negative (FN) - Recognised person's voice incorrectly identified as not-recognised person

The Fig.2 illustrates the performance of employed techniques w.r.t to real-time speaker voice database for recognition accuracy and Sensitivity standard measures. The performance of MMBO association in configuring ANN model demonstrates greater forecasting performance than other strategies implemented in this research, as seen in the following graphical depiction.

Accuracy: Accuracy is also used as a statistical measure to appropriately detect/reject the recognized / not-recognised person with respect to authenticate biometric characteristics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (12)$$

Table 3. Performance measures of employed techniques

Techniques	TP	TN	FP	FN	Accuracy	Sensitivity
MBO-ANN	1.8	37.8	0.2	0.2	0.9900	0.900
OMBO-ANN	1.85	37.85	0.15	0.15	0.9925	0.925
CMBO-ANN	1.9	37.9	0.1	0.1	0.9950	0.950
MMBO-ANN	1.95	37.95	0.05	0.05	0.9975	0.975

Table 3 exhibits the performance measures for real-time from all employed techniques along with TP, TN, FP and FN. The results exhibits that the performance of proposed technique is better than others.

MMBO configured ANN achieves accuracy of 99.75% that is 0.25% greater than CMBO configured ANN and 0.5% greater than OMBO configured MBO. Similarly MMBO configured ANN achieves sensitivity of 97.5% that is 2.5% greater than CMBO configured ANN and 5.0% greater than OMBO configured MBO.

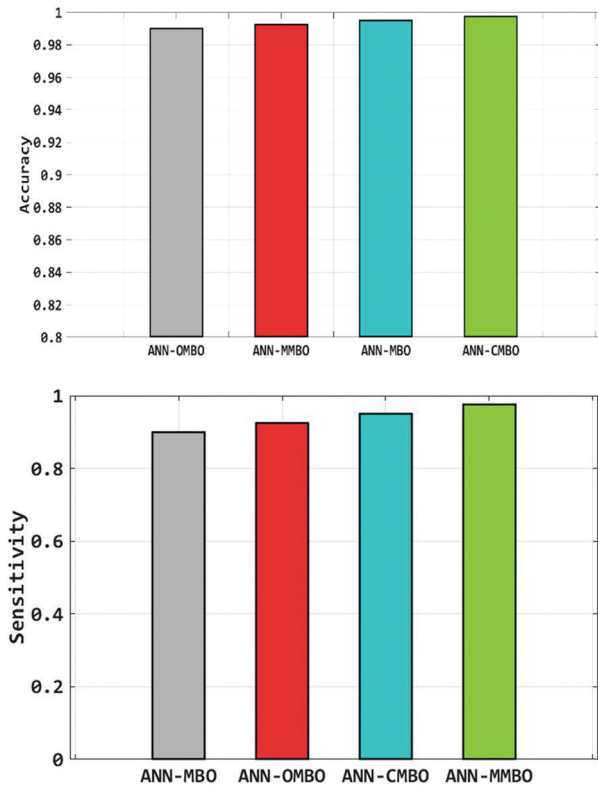


Fig. 3. Performance of Employed Techniques w.r.t standard measures

Converging Performance of the Employed Techniques

The following convergence graph shown in Fig. 4 signifies the performance of optimization techniques integrate with ANN for model configuration in real time speaker voice database. The performance of the used optimization association ANN approaches starts at the same point and gradually becomes exponential up to the 100th iteration, slowing down marginally after that. The proposed MMBO saturates at 400th iteration, which quite early over contest techniques; this is possible because of mutating both opposition and Cd strategy in traditional MBO.

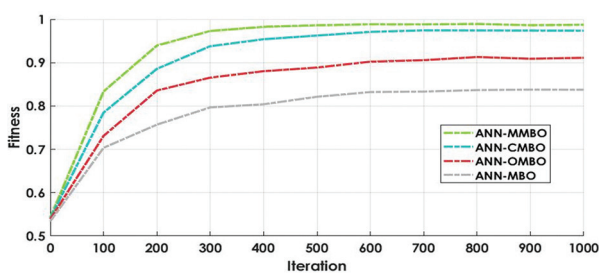


Fig. 4. Convergence graph

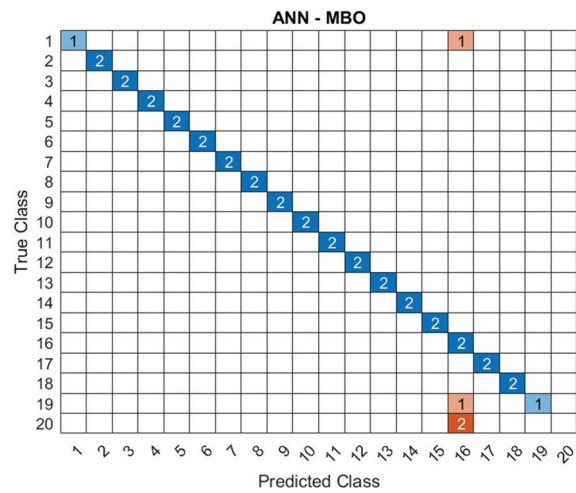


Fig. 5. Confusion matrix for real-time testing database by means of MBO – ANN

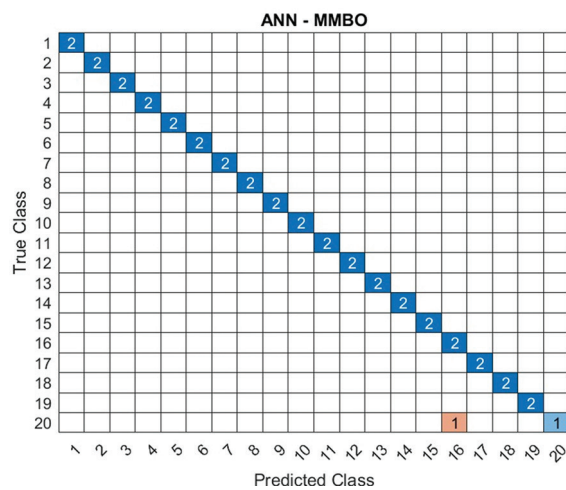


Fig. 6. Confusion matrix for real-time testing database by means of MMBO – ANN

Fig. 5 and Fig. 6 show the confusion charts for real-time speaker recognition using MBO-configured ANN and Mutated MBO-configured ANN, respectively. It is also evident from these figures that accuracy has improved.

5. CONCLUSION

Speaker recognition involves recognizing a person from a spoken word using a machine. It is possible to use speaker recognition systems either to recognize a specific individual or to validate the stated identification of that individual. In this study, voice data are gathered from cooperative office users with no unfavorable microphones. By using MMBO configured ANN models, we were able to recognize the speaker's voice with 97.5% sensitivity, which is superior to contest techniques. The hidden layer identifies characteristics in the input data and uses them to correlate an input with the appropriate output. An increase in hidden layers would complicate the model and lead to overfitting. A manual or trial-and-error approach to achieving those optimal configurations takes time and is time-consuming. Therefore, this research involves integrating opti-

mization techniques and the superior performance is due to the modification of two important strategies, oppositional based solution generation and Cacy distribution in MBO. Research will focus in the future on large-scale speaker identification problems, which are quite challenging.

6. REFERENCES:

- [1] N. N. An, N. Q. Thanh, Y. Liu, "Deep CNNs with self-attention for speaker identification", *IEEE Access*, Vol. 7, 2019, pp. 85327-85337.
- [2] G. S. Morrison, F. H. Sahito, G. Jardine, D. Djokic, S. Clavet, S. Bergths, C. G. Dorny, "INTERPOL survey of the use of speaker identification by law enforcement agencies", *Forensic science international*, Vol. 263, 2016, pp. 92-100.
- [3] M. G. Gomar, "System and method for speaker recognition on mobile devices", U.S. Patent 9,042,867, issued May 26, 2015.
- [4] L. Chen, Y. Liu, W. Xiao, Y. Wang, H. Xie, "Speaker-GAN: Speaker identification with conditional generative adversarial network", *Neurocomputing*, Vol. 418, 2020, pp. 211-220.
- [5] R. Jahangir, Y. W. Teh, N. A. Memon, G. Mujtaba, M. Zareei, U. Ishtiaq, M. Z. Akhtar, I. Ali, "Text-independent speaker identification through feature fusion and deep neural network", *IEEE Access*, Vol. 8, 2020, pp. 32187-32202.
- [6] R. Jahangir, Y. W. Teh, H. F. Nweke, G. Mujtaba, M. Ali Al-Garadi, I. Ali, "Speaker identification through artificial intelligence techniques: A comprehensive review and research challenges", *Expert Systems with Applications*, Vol. 171, 2021, p. 114591.
- [7] S. Tirumala, Sremath, S. R. Shahamiri, A. S. Garhwal, R. Wang, "Speaker identification features extraction methods: A systematic review", *Expert Systems with Applications*, Vol. 90, 2017, pp. 250-271.
- [8] N. Lavan, A. M. Burton, S. K. Scott, C. McGettigan, "Flexible voices: Identity perception from variable vocal signals", *Psychonomic bulletin & review*, Vol. 26, No. 1, 2019, pp. 90-102.
- [9] Z. Qawaqneh, A. A. Mallouh, B. D. Barkana, "Deep neural network framework and transformed MFCCs for speaker's age and gender classification", *Knowledge-Based Systems*, Vol. 115, 2017, pp. 5-14.
- [10] N. Almaadeed, A. Aggoun, A. Amira, "Text-independent speaker identification using vowel formants", *Journal of Signal Processing Systems*, Vol. 82, No. 3, 2016, pp. 345-356.
- [11] Z. Ma, H. Yu, Z.-H. Tan, J. Guo, "Text-independent speaker identification using the histogram transform model", *IEEE Access*, Vol. 4, 2016, pp. 9733-9739.
- [12] J.-Y. Lai, S.-L. Wang, A. Liew, X. Shi, "Visual speaker identification and authentication by joint spatio-temporal sparse coding and hierarchical pooling", *Information Sciences*, Vol. 373, 2016, pp. 219-232.
- [13] Y. H. Jung, S. K. Hong, H. S. Wang, J. H. Han, T. X. Pham, H. Park, J. Kim, S. Kang, C. D. Yoo, K. J. Lee, "Flexible piezoelectric acoustic sensors and machine learning for speech processing", *Advanced Materials*, Vol. 32, No. 35, 2020, p. 1904020.
- [14] R. Chakroun, M. Frikha, "New approach for short utterance speaker identification", *IET Signal Processing*, Vol. 12, No. 7, 2018, pp. 873-880.
- [15] N. Almaadeed, A. Aggoun, A. Amira, "Speaker identification using multimodal neural networks and wavelet analysis", *IET Biometrics*, Vol. 4, No. 1, 2015, pp. 18-28.
- [16] N. Saxena, D. Varshney, "Smart Home Security Solutions using Facial Authentication and Speaker Recognition through Artificial Neural Networks", *International Journal of Cognitive Computing in Engineering*, Vol. 2, 2021, pp. 154-164.
- [17] I. Aljarah, H. Faris, S. Mirjalili, "Optimizing connection weights in neural networks using the whale optimization algorithm", *Soft Computing*, Vol. 22, No. 1, 2018, pp. 1-15.
- [18] D. Devikanniga, R. J. S. Raj, "Classification of osteoporosis by artificial neural network based on monarch butterfly optimisation algorithm", *Healthcare technology letters*, Vol. 5, No. 2, 2018, pp. 70-75.
- [19] P. Soltani, E. Hadavandi, "A monarch butterfly optimization-based neural network simulator for prediction of siro-spun yarn tenacity", *Soft Computing*, Vol. 23, No. 20, 2019, pp. 10521-10535.
- [20] H. Faris, I. Aljarah, S. Mirjalili, "Improved monarch butterfly optimization for unconstrained global search and neural network training", *Applied Intelligence*, Vol. 48, No. 2, 2018, pp. 445-464.

- [21] D. Yang, X. Wang, X. Tian, Y. Zhang, "Improving monarch butterfly optimization through simulated annealing strategy", *Journal of Ambient Intelligence and Humanized Computing*, 2020, pp. 1-12.
- [22] X. Yu, W. Xu, C. Li, "Opposition-based learning grey wolf optimizer for global optimization", *Knowledge-Based Systems*, Vol. 226, 2021, p. 107139.
- [23] S. Mahdavi, S. Rahnamayan, K. Deb, "Opposition based learning: A literature review", *Swarm and Evolutionary Computation*, Vol. 39, 2018, 1-23.
- [24] W. Wang, L. Xu, K. Chau, D. Xu, "Yin-Yang firefly algorithm based on dimensionally Cauchy mutation", *Expert Systems with Applications*, Vol. 150, 2020, p. 113216.
- [25] K. Daqrouq, T. A. Tutunji, "Speaker identification using vowels features through a combined method of formants, wavelets, neural network classifiers", *Applied Soft Computing*, Vol. 27, 2015, pp. 231-239.
- [26] O.S. Faragallah, "Robust noise MKMFCC-SVM automatic speaker identification", *International Journal of Speech Technology*, Vol. 21, No. 2, 2018, pp. 185-192.
- [27] C. Chen, W. Wang, Y. He, J. Han, "A bilevel framework for joint optimization of session compensation and classification for speaker identification", *Digital Signal Processing*, Vol. 89, 2019, pp. 104-115.
- [28] V. de Abreu Campos, D. Guimarães Pedronette, "A framework for speaker retrieval and identification through unsupervised learning", *Computer Speech & Language*, Vol. 58, 2019, pp. 153-174.
- [29] S. Safavi, M. Russell, P. Jančovič, "Automatic speaker, age-group and gender identification from children's speech", *Computer Speech & Language*, Vol. 50, 2018, pp. 141-156.
- [30] K. Devi, J., N. H. Singh, Khelchandra Thongam, "Automatic speaker recognition from speech signals using self-organizing feature map and hybrid neural network", *Microprocessors and Microsystems*, Vol. 79, 2020, p. 103264.
- [31] S. Biswas, S. S. Solanki, "Speaker recognition: an enhanced approach to identify singer voice using neural network", *International Journal of Speech Technology*, Vol. 24, No. 1, 2021, pp. 9-21.
- [32] G. G. Wang, S. Deb, Z. Cui, "Monarch butterfly optimization", *Neural computing and applications*, Vol. 31, No. 7, 2019, 1995-2014.
- [33] F. Chakraborty, P. K. Roy, D. Nandi, "Oppositional elephant herding optimization with dynamic Cauchy mutation for multilevel image thresholding", *Evolutionary Intelligence*, Vol. 12, No. 3, 2019, pp. 445-467.

Multimodal Behavioral Biometric Authentication in Smartphones for Covid-19 Pandemic

Original Scientific Paper

Amitabh Thapliyal

Department of Computer Science and Engineering, Delhi Technological University, Delhi, India
Samsung R&D Institute, Noida
amitabh.t@samsung.com

Om Prakash Verma

Department of Electronics and Communication, Delhi Technological University, Delhi, India
opverma@dce.ac.in

Amioy Kumar

Department of Data Science, Intel Corp, Bangalore, India
amioy.iitd@gmail.com

Abstract – The usage of mobile phones has increased multi-fold in recent decades, mostly because of their utility in most aspects of daily life, such as communications, entertainment, and financial transactions. In use cases where users' information is at risk from imposter attacks, biometrics-based authentication systems such as fingerprint or facial recognition are considered the most trustworthy in comparison to PIN, password, or pattern-based authentication systems in smartphones. Biometrics need to be presented at the time of power-on, they cannot be guessed or attacked through brute force and eliminate the possibility of shoulder surfing. However, fingerprints or facial recognition-based systems in smartphones may not be applicable in a pandemic situation like Covid-19, where hand gloves or face masks are mandatory to protect against unwanted exposure of the body parts. This paper investigates the situations in which fingerprints cannot be utilized due to hand gloves and hence presents an alternative biometric system using the multimodal Touchscreen swipe and Keystroke dynamics pattern. We propose a HandGlove mode of authentication where the system will automatically be triggered to authenticate a user based on Touchscreen swipe and Keystroke dynamics patterns. Our experimental results suggest that the proposed multimodal biometric system can operate with high accuracy. We experiment with different classifiers like Isolation Forest Classifier, SVM, k-NN Classifier, and fuzzy logic classifier with SVM to obtain the best authentication accuracy of 99.55% with 197 users on the Samsung Galaxy S20. We further study the problem of untrained external factors which can impact the user experience of authentication system and propose a model based on fuzzy logic to extend the functionality of the system to improve under novel external effects. In this experiment, we considered the untrained external factor of 'sanitized hands' with which the user tries to authenticate and achieved 93.5% accuracy in this scenario. The proposed multimodal system could be one of the most sought approaches for biometrics-based authentication in smartphones in a COVID-19 pandemic situation.

Keywords: Fuzzy Logic, Keystroke, Multimodal Biometrics, Smartphone, Swipe

1. INTRODUCTION

The last decade has seen many evolutions in smartphones with touch displays, bigger screens, large memory, and processors with high capability. The most powerful and advanced systems for smartphones in this decade are Android and IOS, developed by Google and Apple, respectively. In 2018, the mobile smartphone operating system market share worldwide from

these platforms was 98% with Android (76%) and IOS (22%) [1]. With a report from counterpoint research, there were 1.43 billion smartphones sold in the year 2018. According to a report from Strategy Analytics, major players such as Samsung which sold 291.3 million smartphone units, and Apple sold 215 million smartphones worldwide. Smartphones have a huge impact on people's daily lives and are not limited to calls and messaging. Its utility has increased manifold with the

availability of a huge number of utility applications available for the user, including social networking, entertainment, shopping, and financial transactions. Evolution and advancement in network technology with 5G have opened up several possibilities for streaming and Internet-based applications. While all these smartphones provide convenience and improved use cases, it also brings security and privacy issues for individuals as smartphones process a large amount of private and financial data, which can cause serious loss when it falls into the wrong hands. Therefore, a strong user authentication system that provides user access to smartphones is the most important requirement. Traditional authentication approaches in smartphones, such as swipe, PIN, password, and pattern, are prone to various attacks such as shoulder surfing, guessing attacks, brute force attacks, and dictionary attacks. Shoulder surfing is a very common attack in which the user's password is compromised by peeping into the password entry screen while the actual user types in the password [2]. Biometrics such as the face, fingerprints, voice, and iris are some of the authentication solutions that are the recent trend in Smartphones to provide user access. It utilizes the physiological property of the user that needs to be presented at the time of power-on; hence, it cannot be guessed or attacked through brute force and eliminates the possibility of shoulder surfing. However, fingerprints or facial recognition-based systems in smartphones may not be as applicable in pandemic situations like Covid-19, where hand gloves or face masks are mandatory to protect against unwanted exposure of the body parts. Fig. 1 depicts some of the cases where device operations are required to be performed using hand gloves.

This paper investigates the situations in which fingerprints cannot be utilized due to hand gloves and hence presents an alternative biometric system using the multimodal Touchscreen Swipe and Keystroke dynamics pattern. We propose a HandGlove mode of authentication where the system will automatically be triggered to authenticate a user based on Touchscreen Swipe and Keystroke dynamics patterns. The proposed system incorporates the user's touchscreen swipe and typing patterns as a security layer for authentication to ramp up the total security in the system. We propose the use of a fuzzy network classifier to learn the patterns in this multimodal system to reduce the effects of hand gloves and other external factors in user authentication. Our experimental results suggest that the proposed multimodal biometrics system can operate with high accuracy and the HandGlove mode of authentication has very little or no effect of hand gloves on the accuracy of the authentication system. The proposed multimodal system could be one of the most sought approaches for biometrics-based authentication in smartphones in a COVID-19 pandemic situation. The rest of the paper is organized as follows: Section 2 presents the related work on biometric authentication for smartphones and our proposal, Section 3 presents

the proposed HandGlove mode, Section 4 presents the modules of the proposed multimodal system, and Section 5 depicts the experimental results, and finally, conclusions are discussed in Section 6.

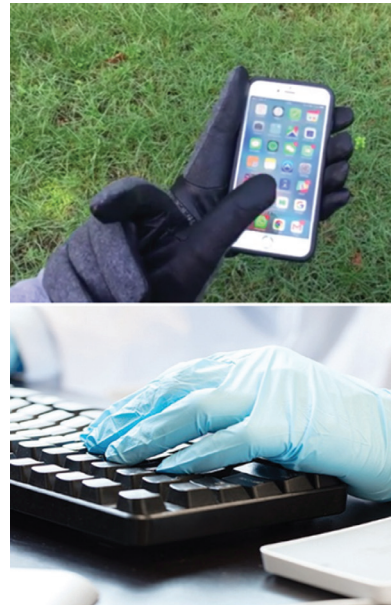


Fig. 1. Device operated using hand gloves

2. RELATED WORK AND OUR PROPOSAL

Several methods have been utilized for authentication purposes to grant users access to smartphones. Some of the popular authentication methods for smartphones are PIN, password, and pattern. However, these methods are not secure, and they have various shortcomings associated with them [3]. Owing to these shortcomings of PIN, password, and pattern-based methods, biometric-based solutions are the recent authentication trends in smartphones. Biometric-based authentication is based on the modalities and traits presented by the user, which can be physical or behavioral patterns of the user based on which they can be recognized by the system.

2.1. RELATED WORK IN SMARTPHONES

If the literature on personal authentication can be arranged chronologically, the biometric traits have been used for authentication for over a century [4], machine-based personal authentication is approximately forty years old [5], and the establishment of automatic biometric authentication as a specific area of research is more than a decade old [6]. In smartphones, the first attempt at bringing a fingerprint sensor was done by Toshiba for their G500 and G900 models in 2007, as shown in Fig. 2. Toshiba used Windows as an operating system in their smartphones, which became instantly popular among people in the days when the current mobile operating systems, Android and iOS, were still not in use. Another smartphone that attempted to

implement a fingerprint sensor was the HTC P6500, which was available in the market after a few months of the G500 release. In 2013, Apple launched Touch ID where the fingerprint was used to unlock smartphones and made available to the iPhone 5S, iPhone 6, iPhone 6 Plus, iPad Air 2, and iPad Mini 3. One of the important security features in Apple Touch ID that makes it very difficult for external imposter attacks is that the fingerprint information is stored locally in a secure location on the Apple chip, instead of being stored remotely on Apple servers or iCloud. The popularity of fingerprint authentication in Apple has paved the way for almost all smartphones to add a fingerprint sensor to their flagships, for example, Galaxy S5/S6, iPhone 5S/6/6S, Huawei Mate S/Ascend, HTC M9+, Xperia Z5, One Plus Two, LG V10, etc.



Fig. 2. Popular Smartphones which initiated using fingerprint sensor Toshiba 6500, HTC P6500, iPhone 6s

To the best of our knowledge, Apple introduced Face-ID using face recognition in iPhone X for the first time in 2017. The popularity of Face-ID has led to various other Android-based smartphones introducing face recognition for user access. However, face recognition also has several limitations, such as low light accuracy, spoofing attacks using photographs, and user inconveniences [7]. Consequently, behavioral biometrics-based methods have also been utilized for user authentication. Researchers have attempted to understand and learn user behavior patterns and how they interact with systems such as keystrokes, touches, and tapping patterns on the device. These behavioral biometric methods provide several benefits over physiological methods, such as behavioral patterns that can be collected continuously and without user knowledge; they do not require any additional hardware sensors to support them. Some of the key works that researchers have attempted on behavioral biometric traits and their accuracy are listed in Table 1.

Keystroke typing pattern-based biometric authentication is based on the fact that each user's typing pattern is unique and consistent. Many approaches to authenticate a device by keystroke biometrics have been utilized in the literature. Clarke and Furnell [8] studied user authentication using keystroke dynamics on mobile devices. In their work, they have used the key typing pattern of 11-digit telephone numbers and 4-digit security PINs to distinguish users. Their models were based on the generalized regression networks with an accuracy of

EERs ranging from 9% to 16%. Sunghoon Park et al., in their paper "Keystroke dynamics-based authentication for mobile devices", achieved an EER of 13% when applying the "Arthematis rhythms with Cues" [9].

Table 1. Behavioral biometric keystroke and touch dynamics

Study	Work Description	Modality	EER
N. L. Clarke et al. [8]	Authentication using keystroke dynamics	keystroke	9% to 16%
Hwang et. al [9]	Arthematis rhythms with Cues	keystroke	13%
Nan Zheng [10]	Tapping patterns	Touch	3.65%
Wang Y. et al [11]	Support Vector Machine	keystroke	8.70%
Meng et al. [12]	Neural Network with PSO	Touch gestures	2.92%
Pin Shen Teh et al [13]	Gaussian, Z-Score, Standard deviation	Touch	8.50%
Ka-Wing Tse et al [14]	RNN	Touch, keystroke	Accuracy 83.9%

Nan Zheng et al. used the union of four features that is pressure, acceleration, time, and size pulled out from smartphone sensors. Experimental results have shown that their verification system achieves accuracy with averaged equal error rates of 3.65% [10]. Meng et al. [12] leveraged touch behavioral patterns from touch gesture data collected from 20 Android phone users for training several classifiers including neural networks. In their work, they also performed optimization of neural networks by using Particle Swarm Optimization (PSO) and achieved an equal error rate of 2.92%. Pin Shen Teh et al. [13] performed an experiment in which data is collected from 150 subjects, and this dataset is shared in three packages of 50 each. In this process, subjects have to enter the same string 10 times, resulting in 20 samples per subject. The timing data and finger touch size features were captured during subject interaction. Three matching functions were used to compute the likelihood of a test sample. These three functions are Gaussian estimation (GE), Z-score (ZS), and standard deviation (SD) drift. FAR and FRR are measured to estimate the accuracy of a biometric authentication system. The Gaussian estimator (GE) gives the lowest EER value in both cases, that is, 8.55% EER when the input string is 4-digit and 5.49% EER when the input string is 16-digit. Ka-Wing Tse [14] evaluated their approach and formulated a dataset of 31 subjects where each subject had to enter a password 50 times. Temporal features, spatial dynamics features, and swipe features were extracted from the dataset. They used the RNN method, and three unique RNNs were implemented and trained separately. The results from each model were fused to obtain the final results. The results indicate that late fusion yields better results than early fusion, and the best result is achieved by spatial features, which were 83.91%.

2.2. OUR PROPOSAL

The increasing popularity of biometrics in smartphones has attracted considerable research work; thus, the literature has shown the number of potential attempts made in this area. However, our literature survey shows the following areas, which are less explored:

- Most of the biometrics utilized in smartphones are physiological, such as fingerprints, iris, face, etc. Some attempts have been made to use behavioral biometrics such as voice and signature, gait, and keystroke. However, these attempts are very few and are currently not well industrialized in smartphones.
- Most of the available work explores a single-modal biometric approach for user authentication in smartphones. Multimodal systems are mostly not considered because of the complexity of the fusion of two different biometric traits in real-time in smartphones. Multimodal refers to systems that can process and relate information from multiple modalities, in our case touchscreen swipe and keystroke typing patterns.
- Most of the available biometrics in smartphones do not consider a pandemic situation like Covid-19, where hand gloves or face masks are mandatory to protect against unwanted exposure of the body parts. In such situations, the acquisition of biometrics from the user itself is a difficult task, which further limits the use of biometric authentication.

In this research, we propose a multimodal-based behavioral biometric system that uses touchscreen swipe and keystroke dynamics patterns to uniquely identify the user and distinguish them from imposters. The highlights of the proposed work are as follows: We propose a behavioral multimodal biometric system with the fusion of the Touchscreen swipe and Keystroke dynamics. The acquisition of these two biometrics is easy and user-friendly, as both of these modalities can be acquired in one action of the hand. Another important highlight of this work is that it investigates the proposed multimodal for situations where hand gloves can be present at hand. We propose a HandGlove mode of authentication where the system will automatically be triggered to authenticate a user based on Touchscreen swipe and Keystroke dynamic patterns. The proposed HandGlove mode will be triggered by the user, and the system will incorporate the user's touchscreen swipe and typing patterns as a security means to authenticate the user.

We develop a fuzzy network classifier to learn the patterns in this multimodal system to reduce the effects of hand gloves and untrained samples in user authentication. Our experimental results suggest that the proposed multimodal biometrics system operates

well with high accuracy and the HandGlove mode of authentication has very little or no effect of hand gloves on the accuracy of the authentication system. We experiment with different classifiers to obtain the best authentication accuracy of 99.55% with 197 users on the Samsung Galaxy S20. The proposed multimodal system could be one of the most sought approaches for biometrics-based authentication in smartphones in a COVID-19 pandemic situation.

A block diagram of the proposed system is presented in Fig. 3. Data collection is the first major module of our proposed system, which is responsible for extracting the keystroke and touch dynamics data from the user's input sample. Details of how the input samples are acquired are explained in Section 4.1. The feature extraction module extracts feature data from the collected sample. In the proposed work, we have used a multimodal approach, and the user sample has features for Touchscreen Swipe and keystroke dynamics. In the training module, a combined feature vector is generated with the touch-swipe and keystroke dynamics and is passed to the feature classifier after being normalized. The fuzzy logic controller unit in the fuzzy classifier is configured to convert a crisp input into a fuzzy value termed fuzzification (explained in Section 4.4). The authentication unit then makes final decisions on accepting genuine users or rejecting imposters.

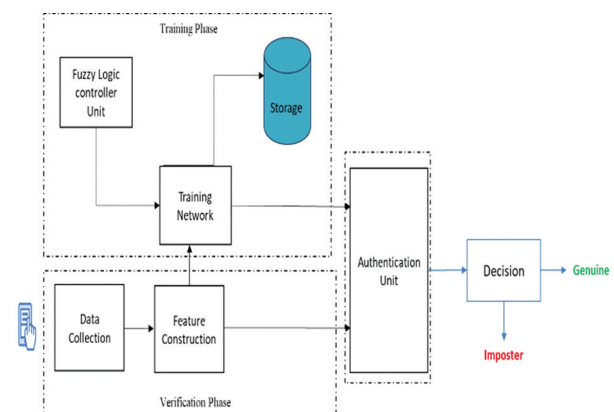


Fig. 3. Multi-Modal Behavioral Authentication Systems

3. HANDGLOVE MODE

The HandGlove mode is used to ease the user. This mode will trigger the multimodal behavioral authentication system and allow device access based on user acceptance by the proposed multimodal system using user swipe and keystroke dynamics. A depiction of the HandGlove mode in mobile devices is shown in Fig. 4.

To detect hand gloves or other external factors on the surface of mobile phones, three main techniques are considered based on the popularity and usability of touch panel devices. Fig. 5a-5c illustrate various detection mechanisms for the detection of hand gloves and other external factors such as wet hands.

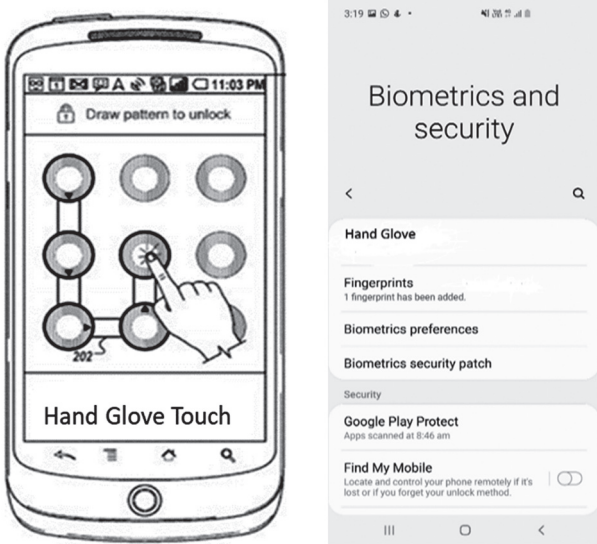


Fig. 4. Hand Glove Mode - Multimodal Behavioral Biometric

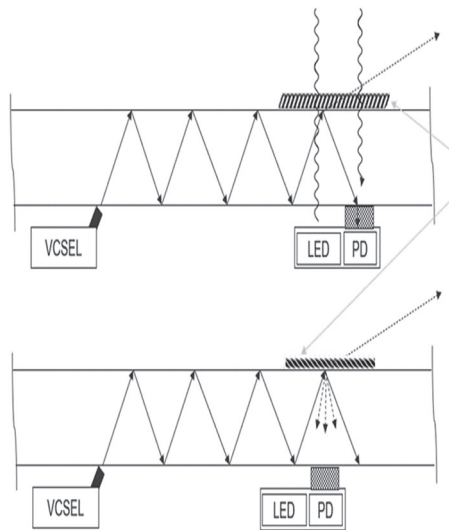


Fig. 5(a)

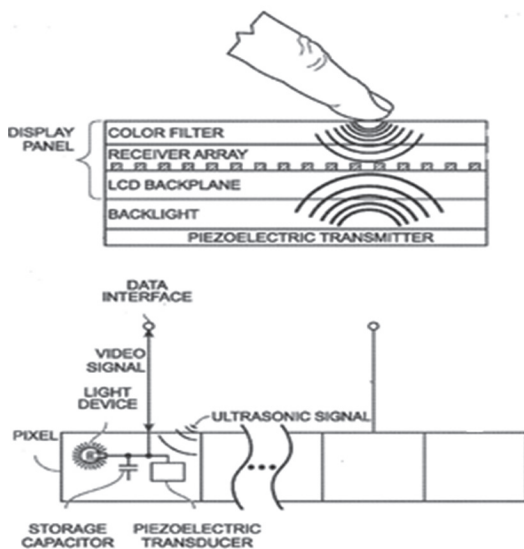


Fig. 5(b)

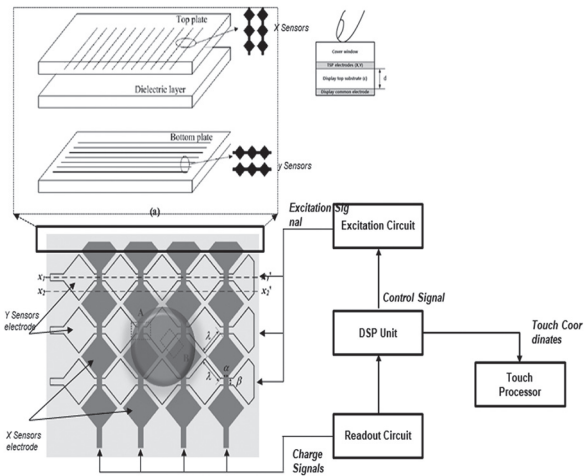


Fig. 5. Detection Mechanism of External Factors
Fig. 5(c)

A detection mechanism based on total internal reflection (TIR) within the display technique is illustrated in Fig. 5a. In physics, TIR is a phenomenon in which the complete reflection of a ray of light within a medium such as water or glass from the surrounding surfaces is reflected into the medium [15]. This phenomenon occurs if the angle of incidence is greater than a certain limiting angle, called the critical angle. Using this principle of total internal reflection, an object along with an external agent is identified. Referring to Fig. 5a, a vertical-cavity surface-emitting laser (VCSEL) or other types of light-emitting diodes capable of producing a controlled beam of infrared light via a lens are provided. When a film/layer of foreign object/contamination/external agent (e.g., finger, gloves, grease, facial oil, water, or other viscous contaminants that may prevent functionality) is present over a proximity sensor (LED), infrared light is reflected into the light detector. As long as the surface is not touched, the light remains inside the screen. However, when an object touches the screen based on the total internal reflection, the light is shattered, so the light escapes from the exact point where the pressure is applied; thus, the position of an object is accurately determined by the sensor registering the light loss. This diversion of the light is also utilized to detect the presence of external agents, such as water, on the surface of the touch screen. The detection mechanism of ultrasonic sensor-based reflection within the display technique is shown in Fig. 5b, which has several advantages over existing technologies for touch screen applications [16].

Finally, the detection mechanism was based on a capacitance-based false positive detection mechanism. Unlike resistive-based touch screens, capacitive screens do not use the pressure of an object to create a change in the flow of electricity. Instead, it works with anything that holds an electrical charge similar to that of human skin. The basic principle of the capacitance-based false-positive detection mechanism is explained

below. As already known, the simplest form of a capacitor consists of two conductors, for example, two metal plates separated by an insulator. The following formula shows the parameters that influence the capacitance:

$$C = \epsilon \times \frac{A}{d} \quad (1)$$

$$\epsilon = \epsilon_0 \times \epsilon_r$$

Where, C is the capacitance, ϵ_r is the relative permittivity (also called the dielectric constant) of the insulating material between the plates, ϵ_0 and is the permittivity of free space (8.854×10^{-12} F/m). A is the area of the plates, and d is the distance between the plates. As shown in Fig. 5c, a flexible and thin display for smart devices having a large coupling capacitance between the sensor electrode of the touch screen panel (TSP) and the display electrode is provided to detect the external agents by utilizing a varying capacitance value that occurs due to the presence of external factors on the touch screen.

4. MODULES OF MULTIMODAL SYSTEM

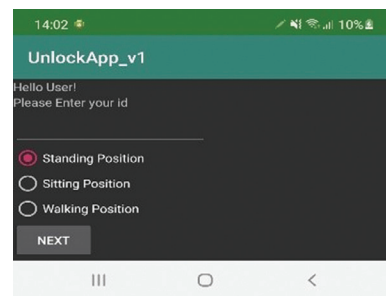
In this section, we discuss in detail the various modules of our system which work together to authenticate the user.

4.1. DATA COLLECTION AND ENROLLMENT

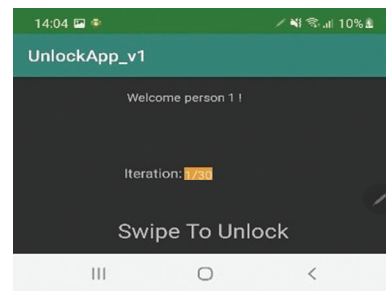
The data collection for the proposed multimodal system can be performed once the android applications trigger the physical sensors to read touchscreen to swipe touch patterns and keystroke password-key input by users. For touchscreen swipe, accelerometer and gyroscope are the sensors used to acquire the user inputs. It captures the touch speed and distance of swipe features corresponding to each enrolled user. For keystroke, we captured the hold-time and inter-key time as a feature for each enrolled user. In contrast to the enrolment module of other biometric systems, the input to the enrolment system in the proposed multimodal system may work in continuous enrolment mode. It can read the above-mentioned features whenever a user swipes and types in a smartphone for better learning of the authentication system. The enrolment system works in the background and reads the swipe pattern and keystroke inputs when the user logs into the system. The application was developed on a Samsung Galaxy S20 device using Google Android OS, 11. We collected data from 197 volunteers (124 men, 73 women) aged between 25 and 40 years. The data collection was done in three different postures: standing, sitting, and walking. The users who participated in the data collection process are presented with a mobile application to collect sensor measurements required to calculate feature values encompassing the behavioral patterns in touch-screen swipe and keystroke dynamics. The users are required to swipe on the application and then type the password appearing on the screen. Each user recorded the data 30 times for each posture

and with three different scenarios of external factors namely dry hands, wet hands, and hands with gloves, making a total of 270 data samples for each user, or 53190 data samples for 197 users. The schematic of the data collection application is presented in Fig. 6.

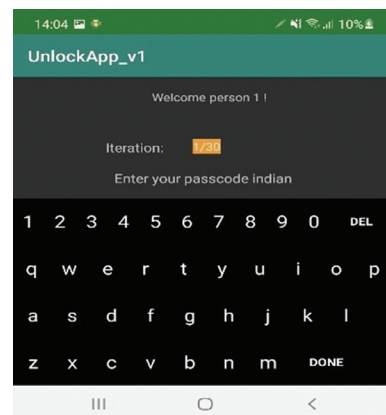
For the experiments reported in this paper, we collected 30 patterns from each individual in each posture. We also asked users to provide inputs with dry hands, wet hands, with gloves as part of data collection, to handle such scenarios to better train the model in HandGlove mode. In total, we collected 53190 samples from 197 users under the three mentioned postures and three external factor cases. Data collection was performed in two separate sessions for each user. The entire enrolment process took 2 weeks period to collect sample data from all 197 users. Data collection and all experiments were performed at the Samsung Research Institute, India R&D, where one of the authors is working. A multimodal spectrogram with data collected considering three scenarios- normal dry hands, gloves, and wet hands—is shown in Fig. 7 for 8 users for better representation.



(a)



(b)



(c)

Fig. 6. Schematic of the Keystroke and Touch-Swipe behavioral data collection application from users. **(a)** Application home-screen where users set their current position. **(b)** Swipe layout where participants are asked to swipe on the screen to capture touch-swipe related feature values. **(c)** Password layout where users type the displayed password on the keyboard to capture keystroke dynamics.

In a preliminary analysis of the data set collected, it was observed that the touch-swipe and keystroke typing patterns of the users when collected have a unique pattern to distinguish the users considering the typing speed (key hold time and key switch time), touch swiping speed, and distance to unlock the device we can identify the smartphone user uniquely.

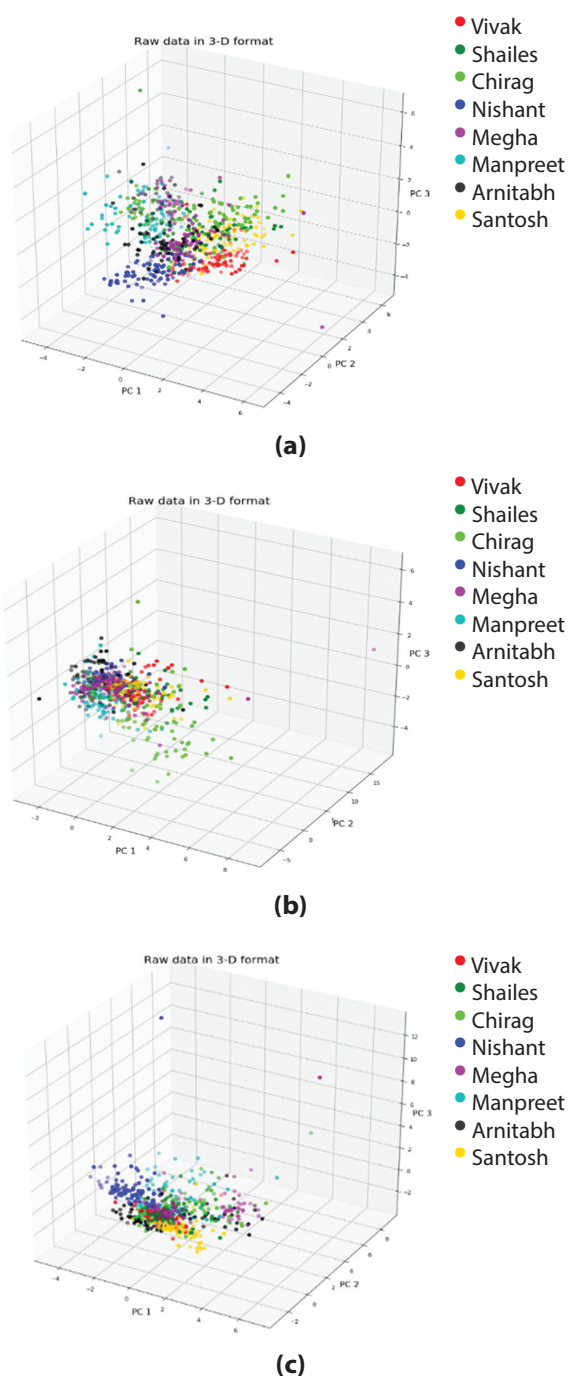


Fig. 7. 3D plot spectrogram of data collected from users with (a) Dry hands, (b) Hand Gloves, and (c) Wet hands

4.2. FEATURE EXTRACTION

The next step is to extract features from the data collected from the user. In the proposed work, we used a multimodal approach, and the user sample has features for Touchscreen Swipe and keystroke dynamics. The feature set captured for Touchscreen Swipe includes the speed of swipe, duration of swipe, and orientation of the touch area with axis along the x-axis when touched on the screen, the orientation of the touch area along the y-axis, accelerometer, and gyroscope. The feature set captured for keystroke dynamics for 6-digit passcode entry-key hold time, key switch time-the time interval to switch from one key to another, also called flight time. A combined vector comprising both modality inputs is the final feature set to be trained with the model. A description of the Touchscreen swipe and keystroke dynamics features is presented in Table 2. To use these features in multimodal authentication, we need to fuse the information extracted from them. Fusion of these features can occur at various levels, such as feature level [17-18], match score level [19], rank level [20], and decision level [21]. The literature work in the areas of biometric authentication has shown that the data fusion at the feature level incorporates the best performance. Hence, in the proposed work, we have utilized the feature level data fusion of the two behavior modalities that is keystroke and Swipe touch features. We combined the feature vectors of the two modalities and generated a combined feature vector with a total of 18 features. However, features extracted from different modalities have different value ranges; therefore, these values should be normalized to represent them in the common range of values.

Table 2. Feature set of the proposed system

Event	Features	Description
Touch	MajorAxis	Orientation of touch area with axis along x-axis when touched on a screen
	MinorAxis	Orientation of touch area with axis along y-axis when touched on a screen
Swipe	SwipeTime	Duration of swipe
	Speed	distance covered by swipe in touch duration
Accelerometer	A_axisMean	mean value of the list of accelerometer values
Gyroscope	G_Jitter	the difference in the ideal signal we type or touch from the gyroscope value
	G_axisMean	axis standard deviation from the list of gyroscope values

Event	Features	Description
Keypad	Key1_Latency	Hold Time Key1
	Key2_Latency	Hold Time Key2
	Key3_Latency	Hold Time Key3
	Key4_Latency	Hold Time Key4
	Key5_Latency	Hold Time Key5
	Key6_Latency	Hold Time Key6
	Key1_2_Latency	key switch time K1->K2
	Key2_3_Latency	key switch time K2->K3
	Key3_4_Latency	key switch time K3->K4
	Key4_5_Latency	key switch time K4->K5
	Key5_6_Latency	key switch time K5->K6

In our work, we utilize the min-max normalization, which maps the minimum of a feature to zero, the maximum to one, and everything else to a decimal between 0 and 1 [22]. Given a set of N feature vectors x_1, x_2, \dots, x_N , we normalize them as

$$x_{ij} = \frac{x_{ij} - x_{min,j}}{x_{max,j} - x_{min,j}} \quad (3)$$

Where, x_{min} and x_{max} are calculated as

$$x_{min,j} = \min_{i=1 \text{ to } N} x_{ij} \\ \& \\ x_{max,j} = \max_{i=1 \text{ to } N} x_{ij}$$

4.3. CLASSIFIER

After the feature extraction step, we experiment with three different classifiers namely

- Isolation Forest (IF)
- k-NN Classifier
- Radial SVM.

We partitioned the dataset into training and test sets and trained these classifiers on the training set. Each model was trained on the combined dataset of the presence of different external factors. The external factors considered in our experiments are dry hands (normal), wet hands (water), and hands with gloves. Each classifier was trained on the combined dataset collected under these three external factors present from each volunteer. The dataset also constitutes key-stroke and swipe dynamics collected under three different positions of the subject, while: Standing, Sitting, and Walking. Evaluation of the model network involves computation of the false acceptance rate (FAR), false rejection rate (FRR), and equal error rate (EER), as discussed in more detail in Section 5.

4.3.1. Isolation Forest

Isolation Forest works on the principle of the decision tree algorithm and is an unsupervised technique mostly utilized for anomaly detection. This algorithm recursively generates partitions on the datasets by randomly selecting a feature and then randomly selecting

a split value for the feature. Anomalies are patterns that have features that are dissimilar to the usual cases. It exploits the fact that anomalous feature observations are few and significantly different from normal observations. Let S be an anomaly score at an instance t .

Then,

$$s(t, m) = 2^{-\frac{E(p(t))}{k(m)}} \quad (4)$$

Where, $p(t)$: length of a point t is computed by the number of edges t covered in the tree until the traversal is terminated.

$k(m)$ is the average of $p(t)$ for specified m

$$k(m) = 2p(m-1) - \frac{2(m-1)}{m}$$

Here, $E(p(t))$ is the mean of $p(t)$ from a group of isolation trees. Using the anomaly score, we can make the following assessments:

- Values close to 1 are considered an anomaly
- smaller than 0.5 considered as normal instances

We split the dataset of samples from each individual into test and training sets and train an individual isolation forest model for each individual. The samples from each person are divided into training and test sets at 85:15 proportion. For the model trained on each individual, we use the rest of the individuals' samples as test samples to evaluate the accuracy of that model.

4.3.2. k - Nearest Neighbor

k-Nearest Neighbor (k-NN) is a simple supervised classification algorithm that can be applied to both classification and regression problems. For each query sample, it finds the k number of nearest samples from the train set in the feature space according to a distance metric. We train a k-NN classifier model on our dataset as a multi-class classification model assigning a label of target identity for the test sample. We divide the entire dataset into training and test sets randomly at 85:15 proportion and classify the test samples and record the FAR, FRR, and EER of the model for evaluation. By tuning the hyper-parameters using the validation set, we used $k=5$ in all our experiments with k-NN. For the distance metric, we used the Minkowski distance metric which is computed as follows.

Let $X=(x_1, x_2, \dots, x_n)$ and $Y=(y_1, y_2, \dots, y_n)$ be the two points in the feature space. Then the Minkowski distance of order p between those two points is given by

$$D(X, Y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p} \quad (5)$$

4.3.3. Radial Support Vector Machine

Support Vector Machines are primarily used for binary classification problems. They simply generate the hyperplanes to separate/classify data in some feature space into different regions. The nonlinearity is added into SVM to work well on high dimensional and linear-

ly inseparable data using a mechanism called Kernel Trick. The Kernel function is of the form

$$K(X, Y) = (1 + \sum_{j=1}^p x_{ij} y_{ij})^d \quad (6)$$

Here d is the degree of the polynomial. In our experiments, we use the Radial Kernel function, which is of the form,

$$K(X, Y) = \exp(-\gamma \sum_{j=1}^p (x_{ij} - y_{ij})^2) \quad (7)$$

Where γ is the hyper-parameter that controls the smoothness of the decision boundary and in turn regularizes the model. The regularization strength of the model is inversely proportional to γ . The SVMs can be used for multi-class classification problems in many different ways. We train the N number of SVM classifiers, where N is the number of identities/classes in the dataset. Each classifier learns the decision boundary between its specific class and the rest of the classes. For a new test sample, we compute the score on each classifier and decide the target class by combining all the scores.

4.4. FUZZY NETWORK

In the proposed work for HandGlove mode, we trained the network for external factors such as dry hands, wet hands, and hand gloves. However, there can be external factors other than hand gloves that could impact user input. For example, the user's hand could be affected by sanitizers, dust, oil or grease, cloth gloves, and so on. This may bring vagueness to the input presented from the user during the authentication phase with HandGlove mode, and we observed high false-positive cases. In such scenarios, the conventional machine learning-based classifiers may not be decisive and fail to handle the test input because their network is not trained for all external factors. To handle such a situation, we train a fuzzy logic classifier with SVM to incorporate fuzziness to minimize the effect of external factors on verification accuracy.

A membership function for a fuzzy set A on the universe of discourse X is defined as $\mu_A: X \rightarrow [0,1]$, where each element of X is mapped to a value between 0 and 1. This value, called membership value or degree of membership, quantifies the grade of the membership of the element in X to the fuzzy set A . Membership functions allow us to graphically represent a fuzzy set. The input to the membership functions are the feature values and the output is the degree of membership in the $[0, 1]$ interval for each fuzzy set. In our experiment, we implement a triangular membership function as shown in Fig. 7.

It contains a lower limit ' a ', upper limit ' b ', and ' m ', where $a < m < b$. In our case, as the feature vector $X=(x_1, x_2, \dots, x_M) \in \mathbb{R}^M$ is the input to the membership function, the parameters a, b , and m are also M -dimensional i.e., $a, b, m \in \mathbb{R}^M$. The membership function is as follows.

$$\mu_A(x_i) = \begin{cases} 0, & x_i \leq a_i \\ \frac{x_i - a_i}{m_i - a_i}, & a_i < x_i < m_i \\ \frac{b_i - x_i}{b_i - m_i}, & m_i < x_i < b_i \\ 0, & x_i \geq b_i \end{cases} \text{ for all } i = 1, 2, \dots, M \quad (8)$$

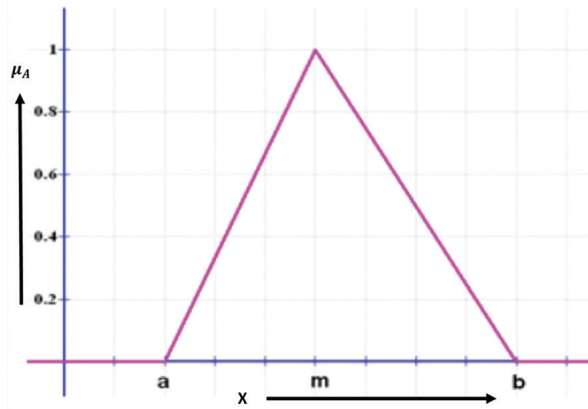


Fig. 8. Triangular membership function

We tune the values of vectors a, m , and b based on the training set to minimize training recognition error. The authentication module makes the authentication decision that the claimant sample matches with the owner of the device. In this case, the claimant user sample features are matched against the stored model, and the degree of membership for each fuzzy set is computed. In the matching process, the degree of membership is compared to the threshold value; if the membership degree is higher than the threshold value, the sample is classified as genuine otherwise impostor. During HandGlove mode in the case of input from the user impacted by the external factor, which is not trained example hands with sanitizer, oil, grease, etc., the proposed fuzzy logic classifier helps to get consistent performance on untrained cases.

5. EXPERIMENTAL RESULTS AND ANALYSIS

This section describes the evaluation method for the proposed multimodal behavioral biometric system on the test data set. The following sub-sections cover the evaluation metrics, methodology, results, and analysis.

5.1. EVALUATION METRICS

The accuracy of the proposed multimodal behavioral biometric system was measured using the following metrics

- False rejection rate (FRR): It is defined as the probability of a genuine user being rejected as an impostor. It is measured as the fraction of the genuine user's score below the predefined threshold.
- False acceptance rate (FAR): FAR is defined as the probability of an impostor being accepted as a genuine user. It is measured as the fraction of the impostor score (a matching score that involves comparing two biometric samples originating from different users) exceeding the predefined threshold.
- Equal error rate (EER): This is used to determine the accuracy of the proposed biometric system.

When both FAR and FRR rates are equal, the intersection point is the EER. The lower the value of EER, the higher the precision of the biometric system.

The relationship between FRR, FAR, and EER is shown in Fig. 9.

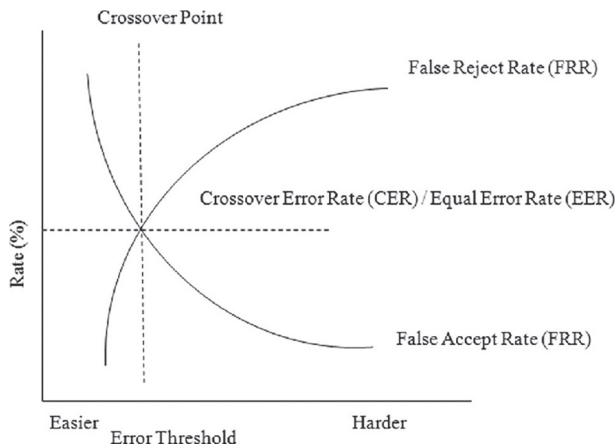


Fig. 9. Relation between FAR and FRR

5.3. EVALUATION METHODOLOGY

To evaluate the accuracy of the proposed multimodal behavioral biometric system based on touchscreen Swipe and keystroke dynamics [23-32], we performed the following task in our experiments to train with binary classifiers such as Isolation Forest, k-NN, SVM, and Fuzzy with SVM Classifier. First, we divided the subjects into two parts: one was treated as the genuine subject and the other as the imposter subject. In our experiment, a total of 197 users participated; for every mobile device, one user is the owner of the device, and his/her samples are labeled as genuine and the remaining 196 users are labeled as imposters. We partitioned the collected dataset into training and test sets in a ratio of 85:15 and trained these classifiers on the training set. We generated four models using four different training sets for different postures: sitting, standing, walking, and all postures. Both the training and the test sets contained all the variations in the external factors (dry hands, wet hands, and hands with gloves). Finally, based on the decision, the evaluation metric values were computed on testing data.

We have four sets of data samples: a genuine training set, a genuine testing set, an imposter training set, and an imposter testing set. Once we have acquired the sample sets, they are used to evaluate the above metrics of the proposed multimodal behavioral biometric system. In the experiments with fuzzy classifier with SVM, users also presented the inputs with non-trained external inputs such as hands with a sanitizer.

For training the k-NN classifier, we simply consider the identities of the users as class labels and train the model as a multi-class classification model. Generally, SVM doesn't support multi-class classification in its

normal form. For multi-class classification, the basic SVM principle is utilized after breaking down the multi-class classification problem into smaller sub-problems, all of which are binary classification problems.

5.3. EVALUATION RESULTS

In experimental results, the EER value was computed for the Isolation Forest Classifier from FAR and FRR values while controlling the 'ease of acceptance' of the isolation forest by varying the contamination factor, and the intersection point in the graph between FAR and FRR for varied contamination level gives us the EER value as shown in the Receiver Operating Characteristic (ROC) Curve plot in Fig. 10.

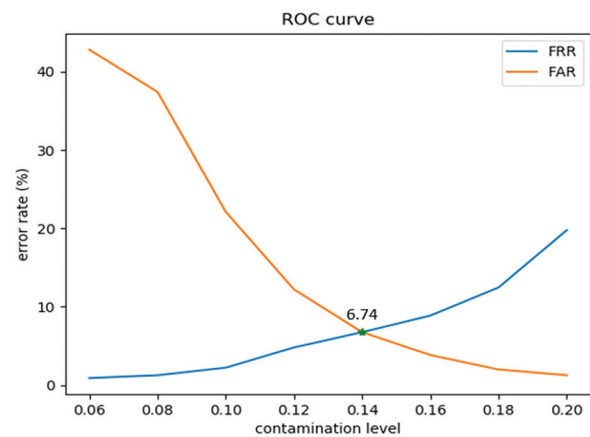


Fig. 10. ROC Curve plot of the proposed system

The equal error rate with isolation forest is obtained at around 6.74% for authentication as shown in Fig. 10. These results are obtained on the combined dataset with and without the presence of external factors such as hand gloves, wet hands, etc. for both training and validation. We conducted experiments by including individual positions in the dataset separately as well as the complete dataset with all three positions standing, sitting and walking. We also experiment with other classifiers such as k-NN and SVM as well and summarize our results in Table 3.

Table 3. Results of proposed Multimodal Behavioral Biometric System with Isolation Forest, k-NN, and SVM classifiers with the test data

Classifier	Posture	Average EER (%)
Isolation Forest	Standing	8.65
	Sitting	6.55
	Walking	8.92
	All	6.74
k-NN	Standing	4.54
	Sitting	4.08
	Walking	4.76
	All	1.58
SVM	Standing	2.04
	Sitting	0.68
	Walking	2.70
	All	0.45

5. 4. ANALYSIS

As per the results mentioned in table 3, we observed that SVM gave the best result of 0.45% equal error rate when including all the positions (Sitting, walking, and standing). SVM is closely followed by k-NN at 1.58% and then isolation forest at 6.74% ERR. The error rates are shown for each posture setting as shown in Fig. 11. Classifiers gave the best results when all the positions are included except for the isolation forest which gave the best result with the ‘Sitting’ position. This shows that the presence of samples of each identity in diverse positions helps to form precise decision boundaries for that identity which further increases the identification accuracy. We note that both touch swipe and keystroke dynamics for all the subjects were considered in the dataset to achieve the results. Further, we observe that the results obtained in the ‘Sitting’ position are better than other positions for all the classifiers as expected because the users are generally more stable while in the sitting position and the variance among the different samples obtained will be minimum. On contrary, the users will be most unstable while walking and so the variance of the samples would be considerably high, and thus walking position accuracy is the lowest.

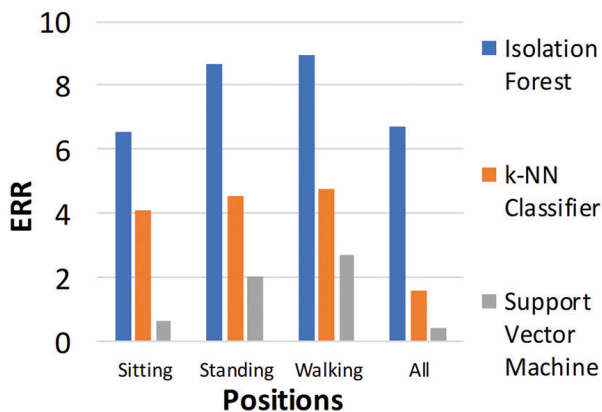


Fig. 11. Chart showing performance of classifiers for each position and all positions together

We also quantitatively compared our work with the recent existing methods utilizing touch-swipe and/or keystroke dynamics behavioral patterns for authentication/verification [8-14] in Table 4. We observed from Table 4, that our approach achieves the Equal Error Rate of 0.45% with the SVM classifier.

However, in a real case scenario, users can try to access the authentication system in presence of varied types of external factors. For example, in the current situation of the pandemic, the user may likely try to authenticate his/her mobile by swiping/typing a pass-code with hands containing sanitizer, dirt or dust, etc. In such cases, the typing or swiping behavioral characteristics may vary slightly due to the presence of such external factors. So, there is a need for a system that can recognize the true owner/ imposter even when the behavioral patterns are slightly varied because of external

factors. Since training the model on the dataset under the influence of all possible external factors is infeasible and impractical, we aim to explore the neighborhood similarities in feature space in an unsupervised manner to solve this problem. We argue that the behavioral features influenced by an unknown external factor ‘a’, will be in near neighborhood space to the behavioral features of ‘closely related’ external factor ‘b’. For example, the behavioral patterns influenced by sanitized hands will be in the near neighborhood to the patterns influenced by wet hands in the feature space because of the closely related physical properties of sanitizer and water. We utilize this contextual neighborhood to train the model to classify into fuzzy sets instead of sharp binary sets such that it incorporates relations between the ‘closely related’ external factors.

Table 4. Behavioral Biometric Keystroke and Touch Dynamics

Study	Work Description	Modality	Average ERR
N. L. Clarke et al. [8]	Authentication using keystroke dynamics	keystroke	9% to 16%
Hwang et. al [9]	Arthematics rhythms with Cues	keystroke	13%
Nan Zheng [10]	Tapping patterns	Touch	3.65%
Wang Y. et al [11]	Support Vector Machine	keystroke	8.70%
Meng et al. [12]	Neural Network with PSO	Touch gestures	2.92%
Pin Shen Teh et al [13]	Gaussian, Z-Score, Standard deviation	Touch	8.50%
Ka-Wing Tse et al [14]	RNN	Touch, keystroke	Accuracy 83.9%
	SVM		0.45%
Proposed work	k-Nearest Neighbor	Touch, Keystroke	1.58%
	Isolation Forest		6.74%
	Fuzzy Classifier		6.50%

For this reason, we train a fuzzy logic classifier on the collected dataset with samples affected by only two external factors namely wet hands and gloves. We considered the triangular membership function as explained in Section 4.4 to decide positive/negative class for a sample and tuned the value of a, b, and m based on the training set. We then utilized the trained fuzzy logic classifier to classify samples of the same individuals affected by an untrained external factor like hands with sanitizer as positive/negative. The results on the untrained external factor are summarized in Table 5. We observe that the error rates of the traditional machine learning-based classifiers increased when tried to evaluate an untrained external factor case. The fuzzy classifier gave the best evaluation results on untrained cases with a 6.46% error rate. This shows that our approach can minimize the effect of external factors like sanitizer, gloves, etc. which are common during the pandemic times like COVID-19 by making use of fuzzy logic.

Table 5. Validation results of the authentication system in the presence of untrained external factor: Hands with sanitizer

Classifier	Average EER (%)
Isolation Forest	22.4
k-NN Classifier	18.25
SVM	16.5
Fuzzy Logic Classifier with SVM	6.46

6. CONCLUSION

This paper investigates the situations in which fingerprints cannot be utilized due to hand gloves and hence presents an alternative biometric system using the multimodal Touchscreen swipe and Keystroke dynamics pattern. We propose a HandGlove mode of authentication where the system will automatically be triggered to authenticate a user based on Touchscreen swipe and Keystroke dynamics patterns. The proposed system incorporates Touchscreen swipe and typing patterns as a security layer for authentication to increase the total security in the system. We propose the use of a fuzzy classification network to incorporate fuzziness in the authentication system with SVM, thereby reducing the effects of unknown external factors such as dust or sanitized hands in user authentication. Our experimental results suggest that the proposed multimodal biometrics system can operate with high accuracy and that the HandGlove mode of authentication has very little or no effect of hand gloves on the accuracy of the authentication system. We experimented with multiple commonly used machine learning-based classification algorithms to obtain the best authentication accuracy of 99.55% with 197 users on the Samsung Galaxy S20. With the developed work accuracy of 99.55% with 197 users with a Samsung Galaxy S20 device and Android R OS, 11.0. The importance of this work is mainly due to the following reasons. First, most of the biometrics utilized in smartphones are physiological, such as fingerprints, iris, face, etc. Some attempts have been made to use behavioral biometrics such as voice and signature, gait, and keystroke. However, these attempts are very few and are currently not well industrialized in smartphones. This work provides a framework for the implementation of a multimodal approach for user authentication in smartphones using touch swipe and keystroke patterns of users. It also provides extensive experimentation on a dataset created using a smartphone (Samsung Galaxy S20). The experimental results established the usability and importance of the presented work for smartphones. We use a fuzzy network to learn the patterns in this multimodal system to reduce the effects of hands with sanitizer in user authentication and achieved 93.5% accuracy on novel external factor case with SVM. The results are shown for 197 users; however, it is sufficient to conclude the potential of the presented work for user authentication in smartphones. More extensive experiments on large

smartphone datasets with more variations in acquisition could be a future scope. To further increase the scope of this work, other modalities such as application usage patterns, battery charging patterns, and walking patterns of an individual can be explored as future research work for smartphone security under a behavioral biometric research scope.

7. REFERENCES

- [1] Counterpoint, "Global Smartphone Market Share: By Quarter", <https://www.counterpointresearch.com/global-smartphone-share/> (accessed: 2022)
- [2] T. Zhao, G. Zhang, L. Zhang, "An Overview of Mobile Devices Security Issues and Countermeasures", Proceedings of the International Conference on Wireless Communication and Sensor Network, Wuhan, China, 13-14 December 2014, pp. 439-443.
- [3] M. Raza, M. Iqbal, M. Sharif, W. Haider, "A survey of password attacks and comparative analysis on methods of secure authentication", World applied sciences Journal, Vol. 19, No. 4, 2012, pp. 439-444.
- [4] T. Sabhanayagam, V. P. Venkatesan, K. Senthama-raikannan, "A Comprehensive Survey on Various Biometric Systems", International Journal of Applied Engineering Research, Vol. 13, No. 5, 2018, pp. 2276-2297.
- [5] N. Ortiz, R. Beleno, R. Moreno, Mauledeoux, O. Sanchez, "Survey of Biometric Pattern Recognition via Machine Learning Techniques", Contemporary Engineering Sciences, Vol. 11, No. 34, 2018, pp. 1677-1694.
- [6] A. Kataria, D. Adhyaru, A. K. Sharma, T. H. Zaveri, "A survey of automated biometric authentication techniques", Proceedings of the Nirma University International Conference on Engineering, Ahmedabad, India, 28-30 November 2013, pp. 1-6.
- [7] S. Ohlyan, S. Sangwan, T. Ahuja, "A Survey On Various Problems & Challenges In Face Recognition", International Journal of Engineering Research & Technology, Vol. 2, No. 6, 2013.
- [8] N. L. Clarke, S. M. Furnell, "Authenticating mobile phone users using keystroke analysis", International Journal of Information Security, 2007, pp. 1-14.
- [9] S. Hwang, S. Cho, S. Park., "Keystroke dynamics-based authentication for mobile devices", Computers & Security, Vol. 28, No. 1-2, 2009, pp. 85-93.

- [10] N. Zheng, K. Bai, H. Huang, H. Wang, "You Are How You Touch: User Verification on smartphones via tapping behaviors", *Proceedings of the IEEE 22nd International Conference on Network Protocols*, Raleigh, NC, USA, 21-24 October 2014, pp. 221-232.
- [11] Y. Wang, C. Wu, K. Zheng, X. Wang, "Improving Reliability: User authentication on smartphones using keystroke biometrics", *IEEE Access*, Vol. 7, 2019, pp. 26218-26228.
- [12] Y. Meng, D. S. Wong, R. Schlegel, L. Kwok, "Touch Gestures Based Biometric Authentication Scheme for Touchscreen Mobile Phones", *Proceedings of Information Security and Cryptology, Lecture Notes in Computer Science*, Vol 7763. Springer, Berlin, Heidelberg.
- [13] P. S. Teh, N. Zhang, A. B. J. Teoh, K. Chen, "Recognizing your touch: Towards strengthening mobile device authentication via touch dynamics integration", *Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia*, December 2015, pp. 108-116.
- [14] K. Tse, K. Hung, "User behavioral biometrics identification on a mobile platform using multimodal fusion of keystroke and swipe dynamics and recurrent neural network", *Proceedings of the IEEE 10th Symposium on Computer Applications & Industrial Electronics*, 2020, pp. 262-267.
- [15] L. Whitehead, M. Mossman, "Reflections on Total Internal Reflection", *Optics and Photonics News*, Vol. 20, No. 2, 2009, pp. 28-34.
- [16] J. Dolcourt, "Galaxy S10 has an ultrasonic fingerprint scanner. Here's why you should care", <https://www.cnet.com/tech/mobile/galaxy-s10-has-ultrasonic-fingerprint-scanner-heres-why-you-should-care-explainer> (accessed: 2022)
- [17] A. Ross, R. Govindarajan, "Feature level fusion using hand and face biometrics", *Proceedings of the SPIE 2nd Conference Biometric Technology Human Identification*, Orlando, FL, USA, 2005, pp. 196-204.
- [18] K. Chang, K. W. Bower, S. Sarkar, B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, 2003, pp. 1160-1165.
- [19] A. Ross, A. K. Jain, "Information fusion in biometrics", *Pattern Recognition Letters*, Vol. 24, No. 13, 2003, pp. 2115-2125.
- [20] A. Ross, K. Nandakumar, A. K. Jain, "Handbook of Multibiometrics", Springer, Boston, MA, 2006, pp. 59-90.
- [21] T. Kinnunen, V. Hautamäki, P. Fränti, "Fusion of spectral feature sets for accurate speaker identification", *Proceedings of the 9th Conference Speech and Computer*, St. Petersburg, Russia, 2004, pp. 361-365.
- [22] S. Gopal, K. Sahu, "Normalization: A Preprocessing Stage", *International Advance Research Journal in Science, Engineering and Technology*, Vol. 2, No. 3, 2015, pp. 20-22.
- [23] J. Kim, H. Kim, P. Kang, "Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection", *Applied Soft Computing*, Vol. 62, 2018, pp. 1077-1087.
- [24] D. Stefan, X. Shu, D. D. Yao, "Robustness of keystroke-dynamics based biometrics against synthetic forgeries", *Computers & Security*, Vol. 31, No. 1, 2012, pp. 109-121.
- [25] A. Motwani, R. Jain, J. Sondhi, "A Multimodal Behavioral Biometric Technique for User Identification using Mouse and Keystroke Dynamics", *International Journal of Computer Applications*, Vol. 111, No. 8, 2015, pp. 15-20.
- [26] V. Stanciu, R. Spolaor, M. Conti, C. Giuffrida, "On the Effectiveness of Sensor-enhanced Keystroke Dynamics Against Statistical Attacks", *Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy*, New Orleans, USA, 9-11 March 2016, pp. 105-112.
- [27] C. Giuffrida, K. Majdanik, M. Conti, H. Bos., "I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics", *Lecture Notes on Computer Science*, Springer, Vol. 8550, 2014, pp. 92-11.
- [28] X. Huang, G. Lund, A. Sapeluk, "Development of a typing behavior recognition mechanism on android", *Proceedings of the IEEE 11th International Conference on Trust, Security and Privacy in Com-*

puting and Communications, Liverpool, UK, 25-27 June 2012, pp. 1342-1347.

- [29] C. J. Tasia, T. Chang, P. C. Cheng, J. H. Lin, "Two novel biometric features in keystroke dynamics authentication systems for touch screen devices", *Security and Communication Networks*, Vol. 7, No. 4, 2014, pp. 750-758.
- [30] D. Umphress, G. Williams, "Identity verification through keyboard characteristics", *International Journal of Man-Machine Studies*, Vol. 23, No. 3, 1985, pp. 263-273.
- [31] R. V. Yampolskiy, V. Govindaraju. "Behavioral Biometrics: A survey and classification", *International Journal of Biometrics*, Vol. 1, No. 1, 2008, pp. 81-113.
- [32] A. Masri, "Active Authentication Using Behavioral Biometrics and Machine Learning," George Mason University, Fairfax, VA, USA, Ph.D. Thesis, 2016.

Decision Support Machine- A Hybrid Model for Sentiment Analysis of News Headlines of Stock Market

Original Scientific Paper

Kirti Sharma

School of Computer Application,
Lovely Professional University, India
kirtisharmaa.11@gmail.com

Rajni Bhalla

School of Computer Application,
Lovely Professional University, India
rajni.12936@lpu.co.in

Abstract – Forecasting and making speculations about the financial market is intriguing and enticing for many of us. Predicting sentiments in the field of finance is a difficult thing as there is a special language that is used in financial markets and the data is unlabeled. Generalized models are not sufficient because the words that are used in financial markets have a completely different meaning when compared to their regular use. This paper represents the study of the stock price fluctuations and forecasting of the future stock prices using financial news about the big IT giants. NLP techniques should be applied to extract the correct sentiments out of the statements. This paper proposes a hybrid Machine Learning model DSM i.e. Decision Support Machine based on Support Vector Machine and Decision Tree. In this study news headlines dataset is preprocessed and then used for making predictions. The results show that the proposed model DSM got an accuracy of 79.75%. Results are then compared with the real-time stock market data for the same time duration, thus giving us a better picture of the actual changes. DSM is also compared with BERT, TextBlob, Decision Tree, Naïve Bayes, NLTK-Vader, SVM and KNN. The proposed model can further be extended if more datasets associated with investors' sentiments can be used for training.

Keywords: Sentiment Analysis, Stock Prediction, TextBlob, NLTK-Vader, BERT, SVM, KNN

1. INTRODUCTION

Finding out the positivity or negativity of sentiments by analyzing the relationship between the sequence of words and building a model on it has become a crucial task. The stock market is highly dynamic. Researchers, Market Professionals, and many other people keep on putting their efforts to find new and accurate ways of finding out the movements of the financial market.

In this study we have used various sentimental analysis techniques on financial news, the focus is only on news headings. Sentiment analysis techniques are used to find out the efficacy of headlines on the financial market.

This study is about finding the impact of financial news headlines related to the NSE IT Giants on stock trends. The Stock Market is very much driven by sentiments of the masses, and government policies, it calls for the application of various Natural Language Pro-

cessing Techniques. In this paper, a hybrid model DSM is proposed which is giving better results in comparison to other sentiment analyses i.e. TextBlob, NLTK-Vader, BERT models along with Naïve Bayes, SVM, K-NN, and Decision Tree models. The main goal is to identify the impact of financial news headlines on the stock market in real time. This study resulted that few of the big news headlines put an impact on the next day's prices.

Data collection forms the basis of research analysis. Dataset is Gathered from reputed and trustworthy financial news based on different IT Giants, followed by the data preparation steps. Various Sentiment Analysis tools are then applied to the dataset. The summary of this study is as follows:

- To identify the effect of a news headline, a hybrid DSM model is developed.
- Two years dataset with news headlines of six IT Giants of NSE is gathered using a python-script, from moneycontrol.com and two years of histor-

ical data is taken from National Stock Exchange's official website

- A comparative study between different models that were used in past on stocks' news and the hybrid proposed model is performed
- Results are plotted to understand the actual effect of the news on the real market.

Many studies have proposed or developed different techniques for identifying the sentiments of news of the financial market. This section highlights some recent work done in this area.

Nemes and Kiss [1] in their paper have shown the impact of news headlines on the stock market by comparing BERT, RNN, and VADER techniques. As per their research, BERT & RNN gave better accuracy without neutral values when compared to other discussed tools. There is a variation in the results of the stock market when various models are applied to the data. RNN model which was trained gave better results in comparison to other SA tools (neutral values are not considered). Kaczmarek and Perez [2] have used random forest on twenty years of stock data to do the monthly forecasting. They have made use of three portfolio-building approaches and made a comparison of these methods - Hierarchical Risk Parity, mean-variance, and 1/N rule, where the first two methods exceed the result over 1/N. Their research has come up with a portfolio-building approach based on RF along with either HRP or mean-variance.

Ingle and Deshmukh [3] used news data for various BSE-listed companies to make forecasting for the next day. Ensemble Deep Learning framework is used for the same. The results gave better accuracy in comparison to other techniques where high, and low values were giving the exact same values. Authors have suggested that high-frequency trading methods in the future can be used to make improvements, but they have not analyzed the effects of the sentiments of the masses.

Khattak, Ali, and Rizvi [4] have analyzed the possible determinants of the European Financial market. They have inspected various possible factors that can influence European market. They have made use of the LASSO technique, their research concluded that Germany and France are the biggest predictors. Devlin et al [5] have developed a new model BERT ("Bidirectional Encoder Representations"). BERT is made in such a way that it uses both left to right and right to left contexts to pretrain the model. Fine-tuning this model is possible just by having one extra output layer. Usmani and Shamsi [6], in their paper, have proposed a news headline classification technique without any training data. They have made use of NLP algorithms to fetch the seed terms, which is then used for refinement of classification outcomes. Data validation is done with ANN and shown with a normalized confusion matrix. They have concluded that when a dataset is classified appropriately, model performance improves signifi-

cantly. Das et al [7] have presented the ways to make economic judgments i.e. Financial market price judgments, and analysed twitter dossier to make predictions about the share prices of a firm. They have used "Spark, Twitter API, Apache Flume", to stream, process and analyse the data.

Lee [8] analyzed early effects of COVID-19 on people's emotions using DNSI and Google trends on corona-related topics about US Financial Market. The aim is to find the interrelation amongst sentiments and eleven chosen US stocks in a said time period. Positivity and negativity of the public's sentiments about financial market crashes can lead to a compound impact on Stockholders. The analysis shows that different companies get affected by variations in the emotions of the public due to Covid-19.

Arras et al.[9] have shown the modest and useful way of using LRP extended method to LSTM approach by laying down the backpropagation axiom. They have made the extension of LRP to a bi-LSTM to make the sentiment forecasting of a given phrase. Using Neural Networks to find the impact of market attributes on the financial market may not do the forecasting accurately, as the randomly chosen problem's weight may lead to inaccurate forecasting. Based on "WORD-VECTOR" in deep learning Pang et al [10] shows the topic of "STOCK VECTOR". They have taken High-Dimensional past data of various stocks as input to the system and have come up with the idea of LSTM with an ingrained layer and auto-encode technique to do the forecasting. L.Lima et al.[11] in their paper have shown better results in forecasting stock prices, they have worked with Support Vector Machine, and have analyzed all public emotional features. They showed the effect of positive tweets is generally positive on the stock market. K.Herng Leong et al. [12] have created an application that classifies the feedback on the basis of positivity, negativity, and neutrality. Key words were determined and analyzed to find the depth of the emotion. This technique helped in enhancing the standards of the services of the healthcare system. J. Chun et al. [13] have come up with a concept of an emotion-based forecasting system and named it ESPS. Their study centered on taking into account the diverse emotions of investors. ESPS is based on DNN and have considered both positive and negative features i.e., joy, anger, sad, scare, shock, etc. Their study concluded that the ESPS's results are better than the models under consideration. Mujahid et al. [14] have analyzed the sentiments of tweets using TF-IDF and WoW, about online learning during covid pandemic. Their study reveals that Random Forest & SVM along with Bow gave good results. Authors have used various Deep learning techniques like CNN, LSTM, Bi-LSTM, CNN-LSTM and topic-modeling to figure out the issues with online learning. Research study done by G. Aditya et al. [15] have studied and implemented a support vector machine on hotel reviews data. Preprocessing, and conversion of the review are done to figure out the +ve or -ve sentiments.

Literature Survey on stock market trend predictions has drawn our attention to some important points. These points are listed below:

- Most of the research work is conducted with a limited amount of dataset
- A low number of news data instances is used, and a reliable news source is required
- In the case of text data, data collected is unlabeled and that's why there is a need to label the data accurately, which should either be done manually or with the help of some NLP technique
- In order to provide solutions to the above-said problems, a new hybrid model DSM is proposed in this study. The model is pre-processed, trained from the scratch, and then applied to the dataset

Ticker	Top	Date	Time
COFORGE	Wait for Nifty to break above 15,370 before going long: Ashish Biswas of CapitalVia Global Research	21 Feb 2021	8.04 am
COFORGE	Is the current consolidation the right time to go long on Coforge?	10 Feb 2021	8.28 am
COFORGE	Buy Coforge; target of Rs 3051: Prabhudas Lilladher	05 Feb 2021	8.57 pm
COFORGE	Neutral Coforge; target of Rs 2690: Motilal Oswal	02 Feb 2021	9.09 pm
COFORGE	COFORGE LTD. Consolidated December 2020 Net Sales at Rs 1,190.60 crore, up 10.92% Y-o-Y	29 Jan 2021	9.22 am
COFORGE	COFORGE LTD. Standalone December 2020 Net Sales at Rs 612.30 crore, up 3.67% Y-o-Y	29 Jan 2021	9.11 am
COFORGE	Wait for Nifty to break above 15,370 before going long: Ashish Biswas of CapitalVia Global Research	21 Feb 2021	8.04 am
COFORGE	Is the current consolidation the right time to go long on Coforge?	10 Feb 2021	8.28 am
COFORGE	A strong quarter from Coforge, but can the stock rally further?	23 Oct 2020	10.14 am
COFORGE	Why should investors hold on to Coforge despite the stupendous rally?	28 Sep 2020	9.27 am
COFORGE	Buy NIIT Technologies; target of Rs 2200: ICICI Direct	03 Aug 2020	11.37 am
COFORGE	Sell NIIT Technologies; target of Rs 1540: Dolat Capital	30 Jul 2020	7.09 pm
COFORGE	Buy NIIT Technologies; target of Rs 2190: Prabhudas Lilladher	30 Jul 2020	6.26 pm
COFORGE	NIIT Technologies Q1 profit dips 21% at Rs 80 crore	28 Jul 2020	8.15 pm
COFORGE	Hot Stocks NIIT Tech, Bajaj Auto, Berger Paints three intraday trading ideas	16 Jul 2020	7.40 am
COFORGE	NIIT Technologies - life beyond buyback	01 Jun 2020	9.15 am

Fig. 1. Part from the economics news headlines data frame

2. THE PROPOSED MODEL AND ITS IMPLEMENTATION

Proposed Model DSM works with the data collected from two sources, first from moneycontrol.com, which is a very reputed and reliable source of financial news, and second the historical data from the official website of NSE. Data is scraped using BeautifulSoup, a python library. The scraping technique was capable enough to fetch the news data for specific symbols for the past two years. This data is then extensively pre-processed with the help of tokenization, data standardization, removal of stop-words and stemming, abbreviation pro-

cessing, and token filtering. The historical data used in this study was available on NSE's official website. This numerical data was also pre-processed with the required normalization and transformation techniques.

The proposed model integrates the feature set of both news and historical dataset. It's joined based on the date. The final feature-set is sentiment, opening price, high, low, and closing price. Then the final step is to do the prediction based on the derived feature set. Data sampling i.e., splitting the dataset into training and testing sets is done. The ratio in which splitting was done is 30.

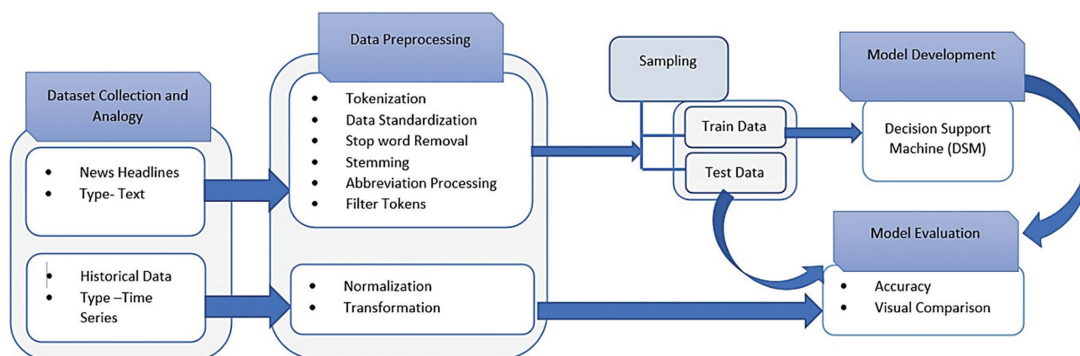


Fig. 2. The overall system architecture of the proposed DSM model

Then hybrid approach for the acquired data set based on the Decision Tree and Support vector machine is applied. Results show that the Decision tree, when applied to our data set gave better classification results for the pos category whereas the Support Vector Machine gave much better accuracy in predicting the neu category. Following these results, when both of these techniques were combined, we achieved much better accuracy. Further, to increase performance, the Tune-sklearn package was used, which is an alternative for Scikit-model Learn's selection module (GridSearchCV, RandomizedSearchCV) with cutting-edge hyperparameter tuning approaches.

The base models for DSM are Support Vector Machine and Decision Tree, which have been optimized further to make predictions with better accuracy. The DSM model is an optimized approach that is trained and tested with well extracted and pre-processed feature set. The accuracy achieved by DSM is 79.75% (Fig. 8.), which is better when compared to the other state-of-the-art models. As shown in Fig. 9., it can be observed that the predictions made by the proposed model for three-class classification achieve better results. It is clear from the table1 that the accuracy of DSM is the highest.

3. METHODOLOGY

3.1. DATA FRAME BUILDING

We have used the financial headlines of the news to do the analysis. Two years of financial News is gathered from 'moneycontrol.com' [16]. Six IT Giants of NSE i.e. COFORGE, HCLTech, Wipro TCS, Tech Mahindra, and Infosys are collected. Dataset is then separated as per the tickers and is stored in CSV. When there is a need to do sentimental analysis, data should be identified as either having positive or negative impacts. As the gathered data was not labeled, part of the data is manually labeled to prepare a labeled training set, which was then used to train the undertaken models. Fig. 1. shows the sample of a raw dataset containing the first few headlines. Part of the news headlines was assigned labels manually to get the best accuracy. Moreover, there was more than one news on the same date, that was also considered and combined based on the date. After that extensive pre-processing i.e. tokenization, standardization, stop-word removal, stemming, and abbreviation processing was done. For historical price data, features i.e. opening price, high, low, the closing price is used. Discretization of this dataset is done. Comparing the values with the previous day's close price, if a value is bigger than the pos, if equal then neu, and in case the value is smaller, the neg value is assigned to the closing price column. Then each row is assigned with a trend value "Up" or "Down", which acts as the input in the proposed model.

The aim of this study is to do a comparison of undertaken companies' stock prices [17]in the pre-decided

period, analyze the sentimental effect of financial news headlines on stock price fluctuations, and determine the actual impact of news headlines. Proposed Model DSM and other sentiment analysis techniques like TextBlob, NLTK-Vader, and BERT, K-NN is used and detailed results are presented in this paper.

3.2. MODELS TRAINED ON DATASETS

Various pre-existing models i.e. TextBlob, NLTK-Vader, Naive Bayes BERT, K-NN, SVM, and DT are trained on the manually labeled and preprocessed news dataset. The models are listed below with descriptions and their accuracy.

3.2.1. TextBlob

[18] Python's API to process the text dossier makes it easier to perform NP tasks. PoS tagging, deriving nouns, categorization, conversion, and a lot more are provided by this API. Depending on the subjectivity, the higher the score is lesser objective the statement is. TextBlob is applied to the dataset and it has shown an accuracy of only 25%. Fig. 3. shows the results of TextBlob on the undertaken dataset. It is visible that the neutral category is the dominating one. Considering INFY, it can be seen that 155 news headings are termed as neutral, 51 positive and only 12 are negative. Fig. 3(b) shows positive, negative, and neutral as 23%, 7%, and 70% respectively. As the neutral category is 70%, the trend is pushing towards positivity, which is leading to distorted results.

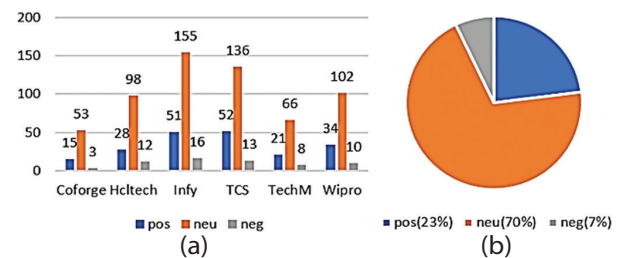


Fig. 3. Company-specific results of the sentiment analysis using TextBlob. (a) Results by Companies and (b) Overall Sentiment Analysis Study

3.2.2. Vader

[19][20] "Valence Aware Dictionary and sentiment Reasoner" is a lexicon and rule-based SA, specially signed to work with sentiments of social networking dossier. Values lie between -1 and +1, negative score means negative statement and positive statement means positive statement. To label the data, the same approach as TextBlob is used. VADER, when applied to the dataset, gave an accuracy of 30% which is again not satisfactory, but a little better than TextBlob which is 25 in this case. Fig. 4. shows that the neutral category is fallen down to 57% and there is an increase in the positive values. This has led to an improvement in the prediction by 5%.

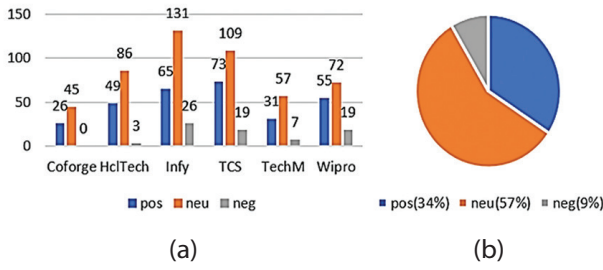


Fig. 4. Company-wise outcome of the sentiment analysis using VADER. (a) Results by Companies and (b) Overall Sentiment Analysis Study

3.2.3. BERT

[5]“Bidirectional Encoder Representations from Transformers” is a pre-trained model in which all words of the statement undertaken are read, and using encoder & decoder models the predictions are made. Text is read by the encoder whereas the decoder makes the predictions. BERT has given an accuracy of 51.36%, which is significantly better than the previous two models. Fig. 5(b) shows that the neu category is reduced to 54%, which is giving a clearer picture of being a sentiment either falls into the pos or neg category. These improved categorizations gave almost double the accuracy of the previous ones.

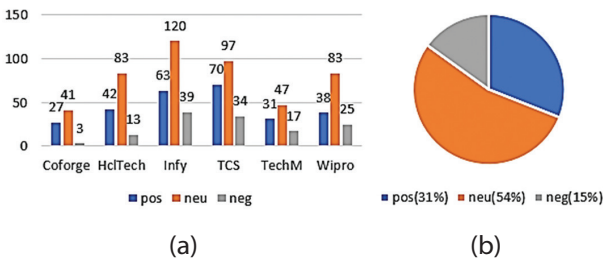


Fig. 5. Company specific results of the sentiment analysis using BERT. (a) Results by Companies and (b) Overall Sentiment Analysis Study

3.2.4. SVM

Multiclass classification is not native to SVM as it does it in a binary way. To achieve multi-classification, there is a need for mapping points in high dimensions. We have used Python and scikit-learn to do this classification.[21] SVM is a supervised learning method and is useful in both classification and regression. [22] It uses a hyperplane to do the classification. When providing the labeled dataset, it outputs an optimal hyperplane that discriminatively classifies the data. It does the prediction based on(1) the pre-defined categories of input. It uses feature space, which is divided by a hyperplane into vectors.

$$\min_{\beta, b, \tau} \frac{1}{2} a^T a + c \sum_{i=1}^n \tau_i \quad (1)$$

$$y_i (a^T \phi(x_i) + b) \geq 1 - \tau_i$$

Where

τ_i =distance to the correct margin with $\tau_i \geq 0, i=1, \dots, n$

c =regularization parameter,

$a^T a = |a|^2$ =normal vector,

$\phi(x_i)$ =transformed input space vector

b =bias parameter, y_i =ith target value

Where vector is given as can be used for high dimensional feature space. [11] SVM is applied to this dataset and it achieves an accuracy of 54.12%.

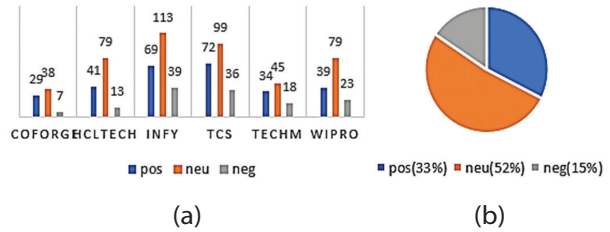


Fig. 6. Company-specific results of the sentiment analysis using SVM. (a) Results by Companies and (b) Overall Sentiment Analysis Study

As shown in Fig. 6., SVM is able to identify the positive category, thereby bringing neutral values down and increasing the possibility of more accurate predictions.

3.2.5. K-NN

K-Nearest Neighbour effectively works where there is a need to classify[23]. It does classification by analogy and computes the equality on the basis of Euclidean Distance. To balance the attributes normalization is applied. More unfamiliar data is given more prevalent class out of its nearest neighbours.

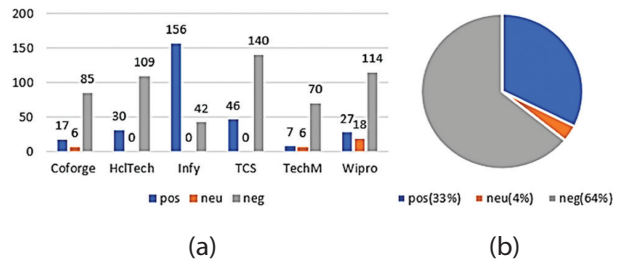


Fig. 7. Company-specific results of the sentiment analysis using K-NN. (a) Results by Companies and (b) Overall Sentiment Analysis Study

After the application of K-NN on the news dataset, we got an accuracy of 57.58%. K-NN has identified a majority of the news as negative but reduced the neutral category values drastically, which when compared to the actual data set is giving better classification.

3.2.6. DT

[14]In the Decision tree, data collection is done in a tree structure or discrete rules. DT has the ability to work with big data. It does splitting based on rules and classifies the data. [24] Recursion results in the addition of new nodes, which is performed till the time criteria

for predicting values are satisfied. The decision tree achieved an accuracy of 61.36%, which can be considered better as it's the highest amongst all the previously mentioned algorithms. Fig. 8. shows how the positive category has substantially increased, which gives the indication of the market moving in a positive direction.

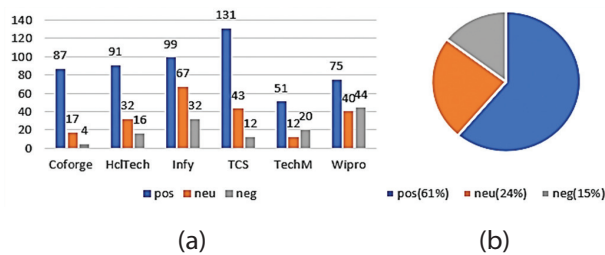


Fig. 8. Company-specific results of the sentiment analysis using DT. (a) Results by Companies and (b) Overall Sentiment Analysis Study

4. RESULTS AND DISCUSSION

This section represents the experimental results to do the predictions of the Stock Market sentiments using DSM and other state-of-the-art models. There are two levels of experimentation, in the first level.

Table 1. Summarised Accuracy Comparison.

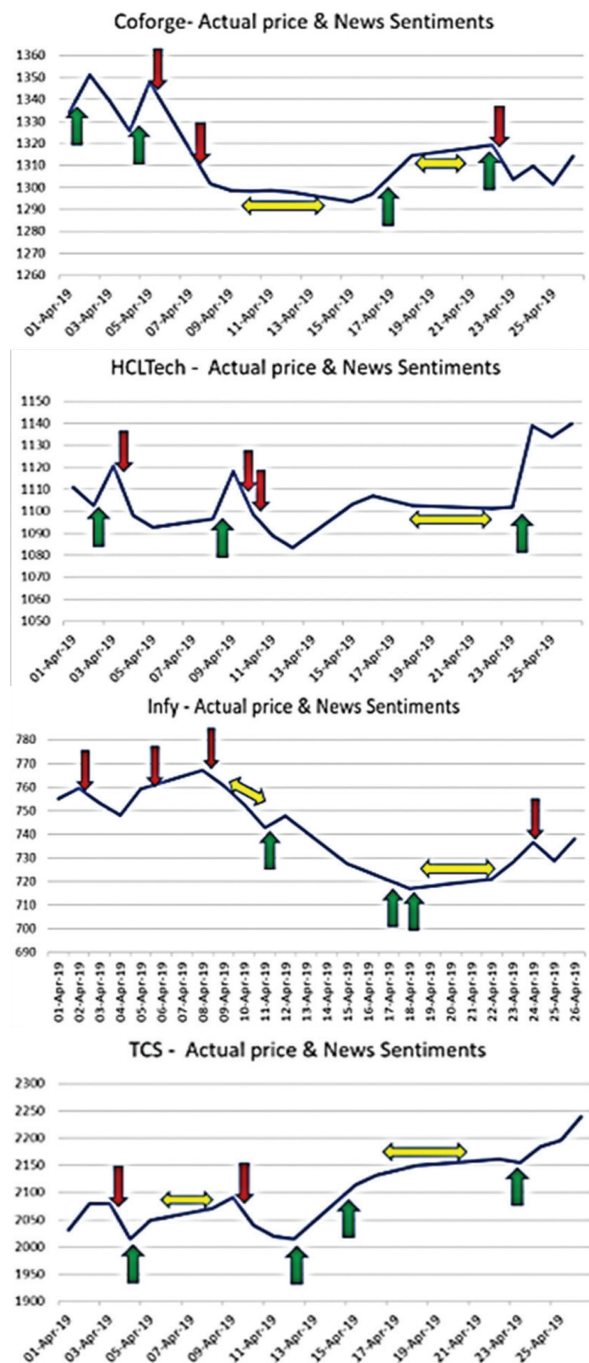
Model Name	Accuracy(%)	Recall	Precision	F1 Score
BERT	51.36	0.142	0.142	0.1428
TextBlob	25	0.193	0.46	0.6365
NLTK-VADER	30	0.453	0.492	0.4469
Decision Tree	61.36	0.438	0.609	0.5748
SVM	54.12	0.372	0.556	0.5193
Naïve Bayes	40.91	0.402	0.325	0.3217
K-NN	57.58	0.513	0.602	0.5546
DSM	79.75	0.585	0.7424	0.6467

Data from six NSE -IT giants have been used for both levels. News headlines with the help of listed techniques are classified and labeled in three classes: pos, neu, and neg. The accuracy in % is displayed in table1. The higher accuracy of the DSM model is achieved in comparison to past studies. Other models have not achieved an accuracy of more than 62%, whereas DSM has reached the level of 79.75% accuracy. Fig. 9. shows the actual stock price along with the derived sentiment from the DSM algorithm. It can be observed that the maximum times the indications given by DSM i.e. calls to buy(↑), sell(↓), or hold(↔) a particular stock is correct except a few times, for example when the actual stock price of Infy was in declining phase, sentiments were not giving clear indications of selling, instead it was showing neutral sentiment that means to hold the stock. But overall, a good clear indication can be found, which can help the investors to take decisions on their holdings.

For the purpose of evaluation of DSM with BERT, TextBlob, NLTK-Vader, Decision Tree, Support Vector Machine, Naïve Bayes, and K-NN, table1 can be referred. The comparison shows that our DSM is outperforming the other models with text data. Further to the combined results, the actual price along with the buy/sell/hold calls indications as represented in Fig. 9. DSM is capable to enhance the trend predictions on the basis of sentiment analysis and historic data.

5. CONCLUSION

Various studies have drawn our attention towards sentimental analysis in today's era, the hybrid approach proposed in this study can be taken as a complementary approach for decision making.



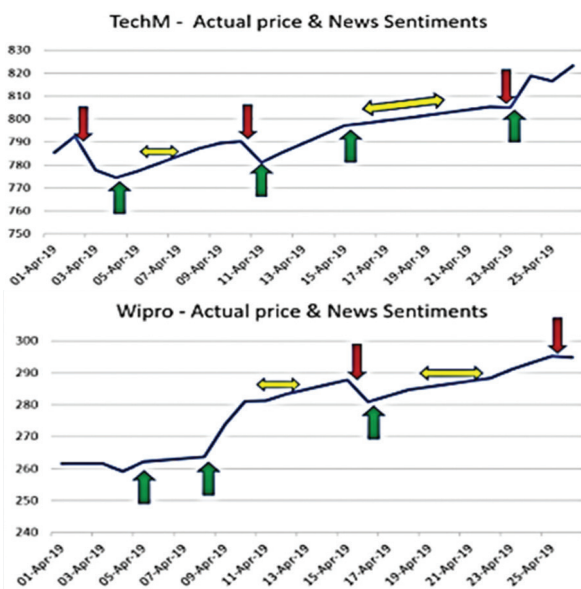


Fig. 9. Buy, sell or hold indications on Coforge, Infy, HclTech, TCS, TechM, and Wipro

The proposed DSM explored the combined effect of news headlines along with historical data in predicting the stock market's trend. The proposed model has shown significant improvement in SM predictions, this study worked on various Sentiment calculation tools to find out the actual sentiment of the undertaken news heading that too without looking into the news details. Sentiments are classified in neg, neu, and pos. The proposed Model DSM works at two levels, first with news headlines and finding polarity and compound scores to classify the headlines in pos, neu and neg. And in the second level, the historical data set is used to find the "Up" and "fall" in the SM for the upcoming day. As shown in Table 1 F1-score of the proposed DSM is 0.6467, which is the best among all the undertaken models. Precision and recall values 0.585 and 0.7424 respectively. Combining both Decision Tree and SVM techniques gave us the advantage of getting better accuracy when it came to pos and neu categories, moreover, the results of neg improved greatly. When applied Tune-sklearn package hyperparameter tuning results improve significantly. The results show that there is a strong co-relation amongst news and SM future trends. These indicators including the full news articles along with annual reports of the companies can also be considered for evaluating the future SM prices.

6. REFERENCES:

[1] L. Nemes, A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines", *Journal of Information and Telecommunication*, 2021, pp. 1-20.

[2] T. Kaczmarek, K. Perez, "Building portfolios based on machine learning predictions", *Economic research - Ekonomska istraživanja*, 2021, pp. 1-19.

[3] V. Ingle, S. Deshmukh, "Ensemble deep learning framework for stock market data prediction (EDLF-DP)", *Global Transitions Proceedings*, Vol. 2, No. 1, 2021, pp. 47-66.

[4] M. A. Khattak, M. Ali, S. A. R. Rizvi, "Predicting the European stock market during COVID-19: A machine learning approach", *MethodsX*, Vol. 8, 2021, p. 101198.

[5] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding", *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 4171-4186.

[10] X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, "An innovative neural network approach for stock market prediction", *Journal of Supercomputing*, Vol. 76, No. 3, 2020, pp. 2098-2118.

[11] M. L.Lima et al. "Using Sentiment Analysis for Stock Exchange Prediction", *International Journal of Artificial Intelligence Applications*, Vol. 7, No. 1, 2016, pp. 59-67.

[13] J. Chun, J. Ahn, Y. Kim, S. Lee, "Using Deep Learning to Develop a Stock Price Prediction Model Based on Individual Investor Emotions", *Journal of Behavioral Finance*, 2020, pp. 1-10.

[14] M. Mujahid et al. "Sentiment analysis and topic modeling on tweets about online education during covid-19", *Applied Sciences*, Vol. 11, No. 18, 2021.

[23] S. Hota, S. Pathak, "KNN classifier based approach for multi-class sentiment analysis of twitter data", *International Journal of Engineering & Technology*, Vol. 7, No. 3, 2018, pp. 1372-1375.

[24] B. Devipriya, Y. Kalpana, "Evaluation of sentiment data using classifier model in rapid miner tool", *International Journal of Engineering and Advanced Technology*, Vol. 9, No. 1, 2019, pp. 2966-2972.

[6] S. Usmani, J. A. Shamsi, "News Headlines Categorization Scheme for Unlabelled Data", 2020. *Int. Conf. Emerg. Trends Smart Technol. ICETST 2020*, 2020.

[7] S. Das, R. K. Behera, M. Kumar, S. K. Rath, "Real-Time Sentiment Analysis of Twitter Streaming data for

- Stock Prediction”, *Procedia Computer Science*, Vol. 132, 2018, pp. 956-964.
- [8] H. S. Lee, “Exploring the initial impact of COVID-19 sentiment on US stock market using big data”, *Sustainability*, Vol. 12, No. 16, 2020.
- [9] L. Arras, G. Montavon, K.-R. Müller, W. Samek, “Explaining Recurrent Neural Network Predictions in Sentiment Analysis”, 2018, pp. 159-168.
- [12] K. H. Leong, D. P. Dahnii, “Classification of Healthcare Service Reviews with Sentiment Analysis to Refine User Satisfaction”, *International Journal of Electrical and Computer Engineering Systems*, Vol. 13, No. 4, 2022, pp. 323-330.
- [15] G. Aditya, T. Priyanka, C. Tanupriya, S. Mohammad, “Sentiment Analysis Using Support Vector Machine”, 2019.
- [20] C. Hutto, E. Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text”, *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, No. 1, 2014, pp. 216-225.
- [21] V. S. Vineeth, H. Kusetogullari, A. Boone, “Forecasting Sales of Truck Components: A Machine Learning Approach”, *Proceedings of the IEEE 10th International Conference on Intelligent Systems*, 2020, pp. 510-516.
- [22] M. H. Krishna, K. Rahamathulla, A. Akbar, “A feature based approach for sentiment analysis using SVM and coreference resolution”, *Proceedings of the International Conference on Inventive Communication and Computational Technologies*, 2017, pp. 397-399.
- [16] “Business News | Stock and Share Market News | Finance News | Sensex Nifty, NSE, BSE Live IPO News.” <https://www.moneycontrol.com/> (accessed: 2022)
- [17] N. S. Exchange, <https://www.nseindia.com/> (accessed: 2022)
- [18] TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation, <https://textblob.readthedocs.io/en/dev/> (accessed: 2022)
- [19] NLTK :: nltk package, <https://www.nltk.org/api/nltk.html> (accessed: 2022)

Identifying and Classifying an Ovarian Cyst using SCBOD (Size and Count-Based Ovarian Detection) Algorithm in Ultrasound Image

Original Scientific Paper

S. Jeevitha

Research Department of Computer Science,
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women,
University of Madras, Chennai, India
jeevisivanandham@gmail.com

N. Priya

Research Department of Computer Science,
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women,
University of Madras, Chennai, India
drnpriya2015@gmail.com

Abstract – Polycystic ovaries are a sign of increasing infertility in the female population worldwide. An excessive number of follicle formations leads to polycystic ovarian syndromes. It affects the female reproductive cycle and leads to disorders such as cardiovascular issues, diabetes mellitus, and cancer. Calculating the number of follicles and detecting the follicle size is still challenging due to time complexity. Since the size of follicles and the greater number mislead the detection of the ovarian type in the ultrasound image. The ultrasound images contain more speckle noise, making the ovarian follicles difficult to see manually. A new convenient method is proposed for the detection of follicles and ovary classification is based on the measurement of size and the count of each follicle. In this paper, the work is divided into four steps, the first step preprocessing the ultrasound image. In the second step, the segmentation process is applied for object selection and separation process using an improved watershed algorithm. In the third step, based on the geometrical and statistical features the object is recognized by SCBOD accurately using various parameters such as size, count, mean, standard deviation, etc., Finally, an SVM classifier is used for classification to conclude the Polycystic ovary syndrome(PCOS) and Non-PCOS. This algorithm is proposed to the physician to find the ovarian follicles rapidly, which will offer accurate performance and is more effective in execution by adopting the SCBOD (Size and Count-based Object Detection) method.

Keywords: SVM Classifier, Polycystic ovary, shape-based Segmentation, size-based Feature Extraction, SCBOD (Size and Count-based Object Detection) method, Improved Watershed Algorithm

1. INTRODUCTION

The ovary is one of the most important reproductive organs in the female reproductive system. It produces an ovum, which consists of follicles in the sac along with some fluids. A dominant follicle will release an oocyte at the time of fertility, when the count of follicles increases several times without the presence of an oocyte, then it is considered a polycystic ovary. The ovary can be classified based on the size and number of follicles shown in the image. Ultrasound images play a crucial role in determining whether that PCOS corresponds to the infertility problem. The ovary can be classified into three types: normal ovary, cystic ovary, and polycystic ovary. **Normal Ovary:** [4] The normal ovary consists of one or two dominant follicles or an-

tral follicles. The size of the follicles is around 2 mm to 28mm and is considered a normal ovary. Antral follicles are said to be less than 18 mm. More than 18 mm below 28 mm are called follicles. **Cyst Ovary:** During the menstrual cycle period, an egg known as a follicle forms inside the sac. The sac opens and produces the egg for the fertility process. If the sac is not open, the fluid in the sac forms a cyst called a cystic ovary. It consists of many types. They are corpus luteum cysts, dermoid cysts, cystadenomas, endometriomas (a type of tissue that forms a cyst ovary), etc. Similarly, some symptoms appear during cyst growth, such as fainting, fever, breast tenderness. The size of the cyst grows around 20mm nearly. **Polycystic Ovary:** The size of the follicle is around 10mm with multiple follicles or collection of fluid increased in the sac that does not release

the egg, then it is considered to be a Polycystic ovary. It can frequently occur due to prolonged menstruation or excessive secretion of a male hormone normally. It results in infertility, type-2 diabetes, and heart disease.

Fig.1,2 & 3 represents the different types of ovaries with different numbers of follicles. The traditional prediction of follicles and classification of ovaries by physicians would be a time-consuming process. To overcome this issue, the proposed SCBOD algorithm will assist the physician to identify and measure the follicles for concluding the PCOS rapidly. The main objective of the proposal is to detect and classify the type of ovaries automatically with accurate results.

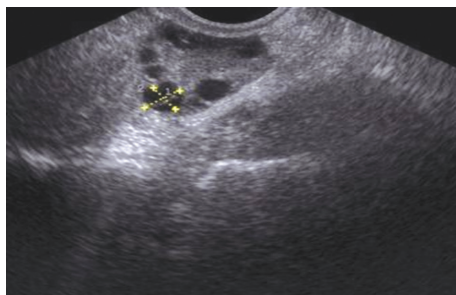


Fig. 1. Normal ovary



Fig. 2. Cystic ovary



Fig. 3. Polycystic ovary

2. RELATED WORK

[1] Aiswarya et al. investigated the feasibility of noise detection in the ovary. It is still a challenge for multiple follicle counts and boundary detection for single follicles. To avoid this kind of problem, a proposal of decomposition of singular values is introduced, along with fuzzy and contour are used for segmentation. [2,22] Hiremath & Froyman examine PCOM (Polycystic ovarian morphology) criteria to detect the PCOS, PCOM

describe the presence of the count and size of the follicle like more than 12 follicles with 2-9mm in size, based on this criteria diagnosis of PCOS concluded. [3] P.S. Hiremath et al, reviewed the geometrical parameter calculation to detect the follicle. The main reason for using minor axis length and major axis length parameters is to estimate the follicle region effortlessly. Gaussian low filtering morphology process, edge-based segmentation help to improve the follicle detection properly in the proposed algorithm [5] Angelos et al. indicated that ovarian number detection could be found in a few seconds using preprocessing techniques and segmentation process with a watershed method. However, these methods are prone to biasing the reporter. Automated concepts would help to make the final decision and increase efficiency, statistical power, and quality by generating the methods. [6] Ranjitha et al, explained the automated system method for finding follicles based on object growing, including different stages. She implemented the speckle noise reduction in the images using median filters, Watershed algorithm is used for follicle extraction. This makes it easier for segmentation sessions. [7] Bhagya et al, reviewed four different approaches to find out the leukemic cell present in the blood. They are edge-based segmentation, canny edge detection, clustering-based segmentation, and morphological watershed-based segmentation. These methods concluded better approaches for performance along with parameters like precision, accuracy, and sensitivity. [8] Eliyani et al, used watershed techniques for follicle detection along with preprocessing techniques for speckle noise removing and follicle segmentation process executed with active contour without edge detection method. [9] Sandy Rihana et al, used preprocessing Morphological techniques for noise reduction, horizontal and vertical line scan threshold implement for further processing, then fusing the binary image for classification to removing the unused area, later the SVM classifier is performed for a classification task. [10] Rose T et al, proposed the follicle monitoring is an essential tool for physicians to diagnose PCOS issues, follicle detection is found using the active contour method with preprocessing techniques. The distance regularized level set used to manage the follicles boundary limits, watershed segmentation also utilized.

[11] Prema T et al proposed the watershed method and Region-based Active contour used for the segmentation process, and classification has been performed based on the geometric features like area, compactness, circularity along with K-NN classifier to detect the follicle in ovary images. [12] Dorin et al used watershed algorithm, active contour model, Gaussian filtering to regularized level set for segmentation process and detect follicle in binary mode ultrasound images. Frechet distance, Error area rate as considered for quality parameters to finding follicles. [13] Mahdi et al investigate regarding shape-based segmentation in three dimensional abdominal ultrasound images using

probabilistic kidney shape based method and confirm the result superiority for segmentation.[14] Hayder et al implemented different classification methods for image classification. The SVM and CNN play a vital role in it. Based on the results concluded the best classifier is the SVM classifier. The results are given according to the image type in this article.[15] Yinhui et al, suggest morphological filtering to preprocess, modify the labeled watershed algorithm for segmentation, a cluster-based method with different criteria was processed for extracting the follicle cysts from the ovary images.

[16] Jun Liu et al, Proposed a fully automated algorithm for object detection using MCL Concepts (Multiple Concentric Layer) which is one of the criteria to find out the focal region of the follicle in the ovary. They processed three different stages as image preprocessing, detection of the focal region of the follicle, and MCL (Multiple Concentric layers); these criteria results are compared with the edge-based segmentation for the follicle detection process.

[17] Carmina et al, Proposed the performance of SVM (Support Vector Method) and CNN (Convolution Neural Network) for tumor detection. The result obtained based on the accuracy and minimal data set feature grade the better result with the SVM method. Since difficulty in finding the follicle's cysts may lead to a problem of efficiency, reliability, and variability. This proposed automated method will overcome all these issues by detecting the follicle in the ultrasound image and counting could be more complicated. In this work, the proposed automated system SCBOD applied the size and count-based detection and separation of the follicle cysts, then the target is extracted by using an improved watershed algorithm. Finally, the classification has been examined using the SVM classifier method.

3. DATASET

The dataset has been collected according to the guidelines of expert doctors from the Athiran Scan's center, and Diagnostics. The total number of images used in the proposed system is 100 images. They are 40 normal ovary images, 40 polycystic ovary images, and 20 cystic ovary images. Distributed dataset for classification as in tab.(1). The pixel size of each image used in the proposed system are 300px by 300px.

Table 1. Dataset Distribution for Classification

Different types of ovary	Training	Testing
Normal ovary(40)	30	10
Cystic(20)	15	5
Polycystic(40)	30	10

4. METHODOLOGY

In the Proposed system SCBOD consist of 4 different phase for identifying the follicle in the ovary ultrasound image, and then ovarian classification was executed based on the different types of ovaries. Some of the criteria used to detect the follicle are the size and count of the follicle with geometrical features. Different phase as follows,

1. Pre-processing
2. Size Based Automated Segmentation process using improved Watershed method.
3. Size and Count Based Object Recognition and Feature Extraction (Object Detection, Separation, and Calculation) using Geometrical features.
4. Classification of Ovary using SVM Classifier.

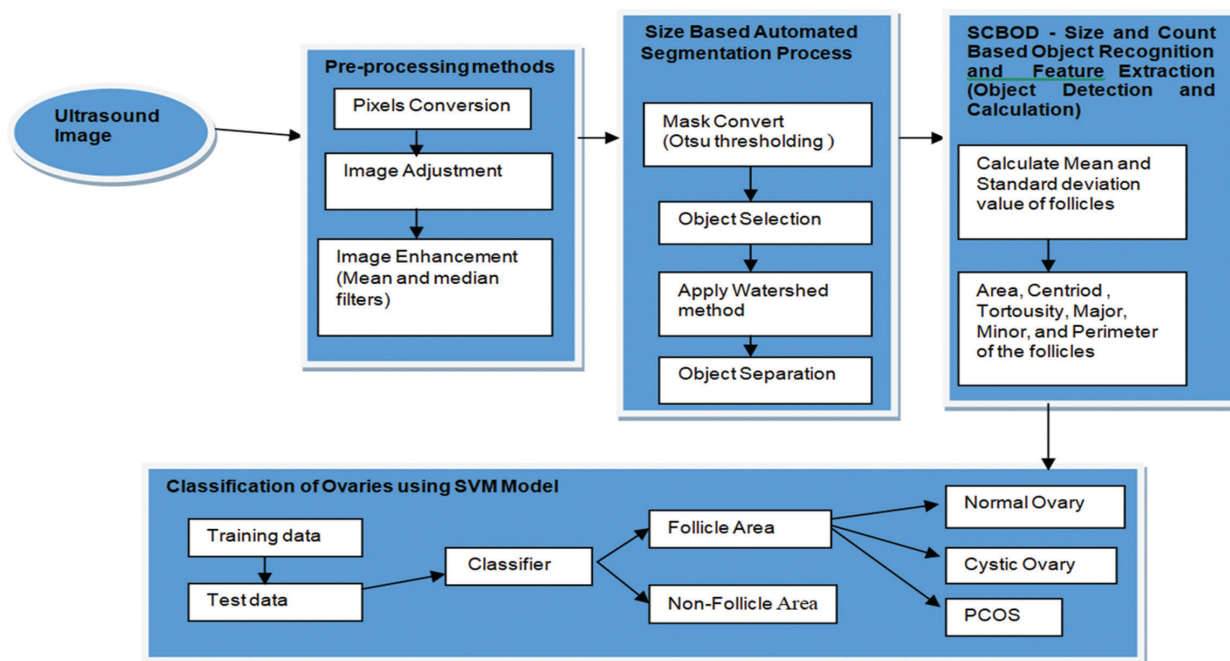


Fig. 4. Framework for the Proposed Algorithms

Fig. 4 represent the proposed algorithm to detect the follicle with preprocessing techniques like Image Adjustment and Image Enhancement, then Segmentation has been done using Mask Convert, object selection done with advanced watershed algorithm and object detection is concluded along with calculating follicle value according to the parameter like Area A, Centroid C, Tortousity Tt, Major Ma, Minor Mi, and Perimeter P. Based on this value finding the follicle region or non-follicle region. Finally, the follicle is classified using an SVM classifier to find out the PCOS or Non-PCOS.

The algorithm for the proposed system is as follows,

Step 1: The pre-processing method has been performed using image pixel conversion, image adjustment, and Image enhancement to reduce the speckle noise reduction process.

Step 2: Size-based Automated segmentation process done by using mask convert, object selection based on shape, normally the ovarian shape is considered oval and sphere, and object separation has been done using the improved watershed method using local maxima.

Step 3: SCBOD algorithm is applied for object detection and feature extraction. Object Detection and Calculation are performed by the size of the follicle. This can be compared based on the ground value of each size of the follicle as, a normal ovary with 8-10 follicles around 2 mm to 28mm in size, cyst 50 to 60 mm(3 to 4 inches), for Polycystic 12 to more follicles exist with 9mm diameter in size. Here the object detection is examined by size and count of follicles using some geometrical features its is shown in tab 1.

Step 4: Finally the SVM classifier is used for the classification process to determine the PCOS or Non-PCOS ovary.

Pseudocode:

Pre-processing method:

1. *To read the input image(L) from the dataset D.*
2. *resize the image by using pixel convert as size_{pix}=(300,300)*
3. *Save the regenerated image after pixel conversion as a new one R.*
4. *then, apply the brightness and Contrast to the saved image.*
If (b!=0) & (b>255) then apply brightnessvalue = 255
5. *finally apply the mean and median filter to remove noise and find the intensities of pixels which are presented in the center of image.*

4.1. SIZE BASED AUTOMATED SEGMENTATION PROCESS USING IMPROVED WATERSHED METHOD

Segmentation is the process of dividing the images or partitioning the image into many regions. Differ-

ent types of segmentation can be done based on the need of image processing. In this phase segmentation, After Applied preprocessing methods to the image and convert the image using mask convert techniques with Otsu threshold value (used for image segmentation also known as binarization). The mask converts the image before the segmentation process to select the object using the edge detection method. Finally Apply the improved watershed algorithms, now some lines appear at the image look as superimposed in the original image Fig. 7. and Fig 8. Fig. 11 represents the object selection and watershed algorithm process.

4.1.1. Mask Convert

Mask Convert implies the inverting image in binary masks whereas white represents 0 and black value 255, Threshold value has to be set automatically according to the image. Otsu threshold is used to reduce the gray level in binary images. After the mask conversion Morphology operation Open, Close, and Dilate, Erode smoothes the objects and removes the small holes and isolated pixels.

4.1.2. Object Selection

During the mask convert all the follicle boundaries close to each other and look like a single object, to overcome this kind of problem. Separating the object is another task in this segmentation process, watershed transformation is used to separate the object, before that edge detection is applied to highlight the sharp difference in the intensity in the selected image, and Maxima is used to find out the segment with one particle be maximum.

4.1.3. Improved Watershed Algorithm and Object Selection and Separation.

The Watershed method is mainly used for object separation from the neighbor images. Watershed lines form at the minimum regions [19]. This is the reason why it's more sensitive with local minimums. To overcome this issue local maximum is implemented (which is opposed to a local minimum) so that it can be separated from the objects along with watershed algorithms. Some of the steps follow in this proposed system by applying watershed algorithms to avoid the over-segmentation issues (finding the edge and applying the watershed algorithm) as follows.

The following pseudo code represents steps involved in the segmentation process.

Pseudocode:

Size and Shape based Automated algorithm:

Mask Convert is used for binarization of image and applied OTSU threshold to detect the optimal threshold value of the pixels.
Set T(threshold value as 0 at beginning), because OTSU threshold method will compute the threshold value automatically.

6. *Input image(L),Ts the size of OSTU threshold,Tos temp object selection, Td object detection*
 1. *for each image (l) to L*
 2. *extrate the set of object region Ro*
 3. *for each object o in Ro*
 4. *if size(o)<Ts*
 5. *remove o from Ro*
 6. *else increase the object selection region to select next.*
3. *After object selection applying the watershed algorithm.*
 1. *After the object selection region, Find the foreground, background region .*
 2. *Generate the local minima of the distance between each object in the foreground. Maxima objects are obtained in the background(opposite).*
 3. *For background used morphological operations and foreground using distance transform.*
4. *Apply improved watershed segmentation algorithm. SU_bg=Surebackground, su_fg=Sureforeground, unk_region=unknowregion, pointer=unknown region*
 1. *Background extraction using morphological operation(dilation and opening for detecting the su_bg)*
 2. *Foreground extraction using distance transformation maximum(for detecting su_bg).*
 3. *Finding the unknown area(unk_region) by subtracting the su_bg and su_fg .*
 4. *Apply watershed algorithm by marking the su_bg != 1 and unk_region =0*
 5. *If (su_bg != 1)*
 - then pointer= pointer + 1*
 - else*
 - if (unk_region=0)*
 - then pointer[unk_region == 255] =0*
 - if (b_region = -1)*
 - then inputimage(pointer == -1) = (255,0,0)*
 6. *The output image with watershed segmentation.(represented in fig.11).*

Fig.11 represents the object selection and improved watershed algorithm process using Regional maxima by opening and closing construction filters.(which is used more efficiently and effectively as normal standard open and close operation. This can give the overall shape of image without disturbance.

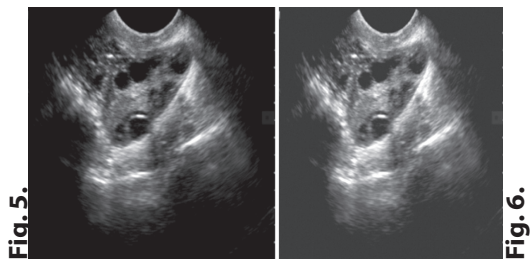


Fig. 5. Bit Conversion
Fig. 6. Image Enhancement

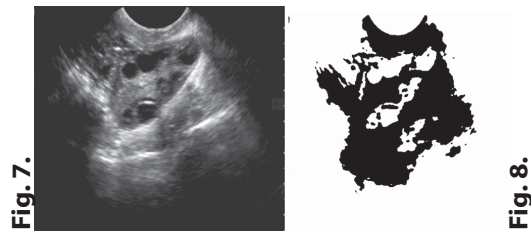


Fig. 7. Original image
Fig. 8. Mask Convert

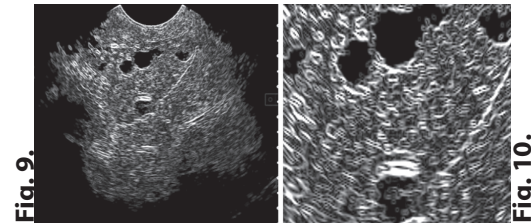


Fig. 9. Object Detection
Fig. 10. Maximum object with edge

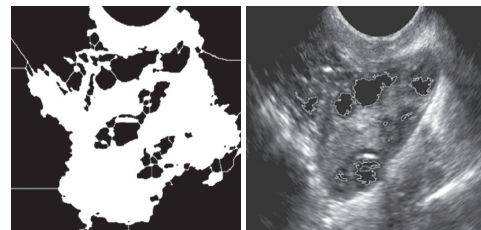


Fig. 11. Object selection using improved Watershed object separation

4.2. SIZE AND COUNT BASED OBJECT RECOGNITION AND FEATURE EXTRACTION(OBJECT DETECTION SEPARATION AND CALCULATION) USING GEOMETRICAL FEATURES

The geometric and statistical operations play an important role in image processing by measuring the objects in ultrasound images. Identification of cysts is diagnosed by calculating the size of follicles under certain limitations [9]. Measuring the follicles is the major concept for extraction. It is required regularly to examine the analysis. Some of the geometrical parameters are used to extract the objects like Area A, Centroid C, Tortuosity Tt, Major and Minor Ma, and Perimeter P, and calculate the parameters of interest to create a new database with this data which is explained in the Table 3.

The parameters used to measure the cyst are in the following manner.

- *Area A is used* to calculate the number of pixels inside the most possible follicles.
- *Major-axis length(a, b),* gives the pixel distance between the major-axis endpoint of the segmented area.

$$\text{Major-axis length}(Ma) = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2} \quad (1)$$

- **Minor-axis length (a, b)**, gives the pixel distance between the minor-axis endpoint of segmented area.

$$\text{Minor-axis length}(Ma) = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2} \quad (2)$$

The major and minor axis represents the ellipse shape at the central moment of the follicle segmented area, because all the cysts or follicle in the shape of oval, its around elliptic one.

- **Centroid C**, is used to measure the center of the object in the follicle ROI(region of interest) it estimates ROI measure by using the a and b coordinates.

$$\text{Centroid } C = (C_x, C_y) \quad (3)$$

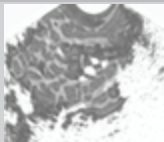


- **Tortuosity Tt** is used to estimate the arc-chord which will give the ratio of curve length(C) and its distance of end length (L) of the estimated follicles.

$$\text{Tortuosity } Tt = C/L \quad (4)$$

- Perimeter is considered as several pixels in the boundary of the follicles. suppose if (a_1, \dots, a_n) is the boundary list, then

$$\text{Perimeter } P = \sum_{j=1}^{n-1} d_j = \sum_{j=1}^{n-1} |a_j - a_{j+1}| \quad (5)$$

Table 2. Standard follicle size and count of ovaries based on OM(Ovary Morphology) and optimized segmentation output along with the count of follicle detection

Different Ovary types	Size of follicles in px(pixel)	Count of follicles	Optimized segmentation image for a different ovary with different counts and sizes of the follicle	Number of follicle present and detected
Normal Ovary	15px-10000px	1-18		Presented 5 Predicated 4
Cystic Ovary	4200px-75000px	1-2		Presented 1 Predicated 1
PCOS	15px-9000px	12-20		Presented 11 Predicated 10

4.3. CLASSIFICATION OF THE OVARY USING SVM CLASSIFIER

Here the main objective for using the geometrical features is to estimate the calculation of the follicle region because the follicle shapes under circular shape or oval in shapes. Based on this feature the dataset is divided into two types are training and testing datasets. The Mean value(Area, Centroid, Tortuosity Tt, Major, and Minor, and Perimeter, and the Standard Deviation values(Area, Centroid C, Tortuosity Tt, Major

Ma, Minor Mi, and Perimeter σP) are utilized for detecting the follicle region based on the classification rules PCOM(Polycystic Ovarian Morphology).

4.3.1 Training Phase and Testing phase

In the training phase, the proposed geometrical feature to calculate the size and count of the follicles using the parameter as Area A, Centroid C, Tortuosity Tt, Major Ma, Minor Mi, and Perimeter P for finding follicle region during the ultrasound image based on the medical expert suggestion and saved for the reference for benchmark calculation. In the testing phase, applied all the geometrical features in the input image to find the segment follicle region and implemented the SVM classification for finding follicle region or non-follicle region.

An SVM classifier is used to classify the ovary into 3 main categories: normal ovary, cystic ovary, and polycystic ovary also compared with the benchmark data for proper classification results. The standard size and count of ovaries fall under the following norms as shown in tab.1.[20]The SVM is the most efficient method for supervised classification models including many probabilistic classifiers. SVM is more popular and easily scalable for a large data set. During the testing phase, these are the parameter that plays a vital role to determine the follicle size and count in the ovarian image and SVM classifier implemented to examine the ovary types as normal ovary, cystic ovary, and polycystic ovary.[21] SVM kernel function is the most important function used to implement the input vector to higher dimensional feature spaces. Polynomial kernel used in the proposed algorithm, which is commonly used by kernels in SVM, and this can satisfy Mercer's condition that can be used for space theory of random processes.

The performance of statistical information of image classification using SVM classifier is listed in Table 2. By precision value, it can be determined the predicate value is accurate. If the precision value of false positive pixels are high. This will give the ratio of true positive to the sum of true positive and false positive. The recall will show all correctly classified instances in the testing phase. F-measure combines both properties of precision and recall as a single measure. These are the parameters used for evaluating the performance of the algorithm. Table 3 shows the confusion matrix for testing performance. While testing phase 39 images are classified as PCOS correctly, 58 images classified as Non-PCOS correctly, only one image as classified incorrectly, this will give 2% wrong prediction from the entire dataset. Total accuracy prediction is 94% from the entire dataset.

5. RESULTS AND DISCUSSIONS:

100 images have been used to determine the ovary types, in that 75 images phase used training and 25 images used for testing phase. Ten-fold cross-validation is used for the testing phase to build a model for the

testing process. based on the size and count of follicle classification of the ovary is examined. In 100 instances there are 40 polycystic ovaries, 20 cystic ovaries, and 40 normal ovaries used, the SVM classifier classified 94% accuracy. Cross verification of accuracy is also executed using the confusion matrix referred to in Tab.6. Similarly Table 5 represents the summary value for the classification validation. Since the database has fewer images will predicate this above accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F - Measure = \frac{2*(Recall*Precision)}{(Recall+Precision)} \quad (9)$$

$$MCC = \frac{(TP*TN)-(FN*FP)}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}} \quad (10)$$

Table 3. Automated measurement of Different follicles with Geometric Feature Extraction.

Geometrical Feature Extraction	Normal ovary	Cystic ovary	PCOS
Area	27	88	42
Perimeter	26.401	87.367	40.792
Major Axis Length	1.128	2.257	3.169
Minor Axis Length	1.128	1.128	6.653
Mean	48.881	6.653	32.498
Standard deviation	22.542	10.113	9.754

In the future, planning to collect different datasets to get better execution and accuracy levels and [4] include some standard error metric features like dice and jaccard for maintain the time complexity. To evaluate the performance of classification Accuracy, Precision, Recall and M-Measure are implemented. they are as follows in Table 4.

Table 4. Performance of SVM classifier during the testing phase

Class	TP Rate	FP Rate	Precision	Recall	F-Measure
PCOS	0.950	0.000	1.000	0.950	0.974
CYST	0.850	0.038	0.909	0.850	0.850
Normal	0.975	0.050	1.000	0.975	0.951
Weighted Avg	0.940	0.028	0.985	0.940	0.940

Table 5. Summary for Cross-validation during Classification of Ovary

Method	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error	Accuracy
SVM Classifier	0.2378%	0.2994%	55.6544%	64.8167%	94%

Table 6. Confusion Matrix of Classification during testing phase

Ovary type	A(PCOS)	B(CYST)	C(NORMAL)
PCOS	38	2	0
CYST	0	17	3
NORMAL	0	1	39

6. CONCLUSION AND FUTURE WORK:

In this paper, the proposed SCBOD algorithm is used to implement the optimized segmentation, feature extraction, and classification of the ultrasound ovary images. The detection and selection are made by the size and count of the follicle, which falls within the ovary image. Based on essential geometrical features and SVM classifier the classification done, the ovary is classified as normal, cystic, and polycystic. The proposed method is capable of detecting PCOS in a short time which can reduce the burden of the physicians to determine PCOS detection difficulty. The proposed method improved the accuracy results by obtaining 94% with good efficiency. Still, there is more confusion about finding the normal cyst and other cysts like endometrioma cysts. In the future, we will keep improving the work to determine the ovary types and give more accuracy to the classification of the ovary, which can be executed by increasing the dataset size for better results and complexity.

7. REFERENCES:

- [1] N. J. Aiswarya, "Automated Follicle Detection in Ultrasound Image of Ovaries", International Journal of Engineering and Technology, Vol. 6, 2019.
- [2] W. Froyman, D. Van Schoubroeck, D. Timmerman, "Automated follicle count using three-dimensional ultrasound in polycystic ovarian morphology", Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 2017.
- [3] P. S. H. Jyothi, R. Tegnoor, "Automatic Detection of Follicles in Ultrasound Images of Ovaries using Edge Based Method", IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition", 2010.
- [4] D. N. H. Thanh, L. T. Thanh, U. Erkan, A. Khamparia, V. B. S. Prasath, "Dermoscopic image segmentation method based on convolution neural networks", International Journal of Computer Applications in Technology, Vol 66, 2021.
- [5] A. Skodras, S. Giannarou, M. Fenwick, S. Franks, J. Stark, K. Hardy, "Object Recognition in the Ovary: Quantification of Oocytes from Microscopic Im-

- ages", Proceedings of the 16th International Conference on Digital Signal Processing, Santorini, Greece, 5-7 July 2009.
- [6] R. Sitheswaran, S. Malarkhodi, "An Effective Automated System in Follicle Identification for Polycystic Ovary Syndrome Using Ultrasound Images", Proceedings of the International Conference on Electronics and Communication System, Coimbatore, India, 13-14 February 2014.
- [7] T. Bhagya, K. Anand, D. S. Kanchana, R. Ajai, "Analysis of Image Segmentation Algorithms for the Effective Detection of Leukemic Cells", Proceedings of Third International Conference on Trends in Electronic and informatics, Tirunelveli, India, 23-25 April 2019.
- [8] S. H. Eliyani, A. Musdholifah, D. Dasuki, "Active Contour Without Edge and Watershed for Follicle Detection in Ultrasound Image of Ovary", Proceedings of the International Conference on Advanced Computer Science and Information Systems, Depok, Indonesia, 17-18 October 2020.
- [9] S. Rihana, H. Moussallen, C. Skaf, C. Yaacoub, "Automated Algorithm for Ovarian Cysts Detection in Ultrasonogram", Proceedings of the 2nd International Conference on Advances in Biomedical Engineering, 2013.
- [10] T. F. Rose, A. K. Styer, E. N. Brown, "Automated Ovarian Follicular Monitoring: A Novel Real-Time Approach", Jeju, Korea, 11-15 July 2017.
- [11] T. Prema, C. Girijamma, "Detection of Cysts in Medical Ultrasound Images of Ovary", Proceedings of the 5th SARC-IRF International Conference, Bangalore, India, 4 May, 2014.
- [12] D. Bibicu, L. Moraru, M. Stratulat, "Diagnostic Accuracy of Ovarian Cyst Segmentation in B-mode Ultrasound Images", Proceedings of the TIM 2012 Physics Conference, 2013.
- [13] M. Marsousi, K. N. Plataniotis, S. Stergiopoulos, "Shape-Based Kidney Detection and Segmentation in Three-Dimensional Abdominal Ultrasound Images", Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26-30 August 2014.
- [14] H. Hasan, H. Z. M. Shafri, M. Habshi, "Comparison Between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models For Hyperspectral Image Classification", Proceedings of the Sustainable Civil and Construction Engineering Conference, 2019.
- [15] Y. Deng, Y. Wang, P. Chen, "Automated detection of Polycystic Ovary Syndrome from ultrasound images", Proceedings of the 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, 20-24 August 2008.
- [16] J. Liu, H. Chen, "Automatic Detection of Follicle in Ultrasound Images of Cattle Ovarian using MCL Method", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetic, 9-12 October 2016.
- [17] C. D. L. Nascimento, S. D. de Souza Silva, T. A. da Silva, W. C. de A. Pereira, M. F. Costa, C. F. F. C. Filho, "Breast tumor classification in ultrasound images using support vector machines and neural networks", Research on BioMedical Engineering, Vol. 32, No. 3, 2016, pp. 283-292.
- [18] X. Ji, Y. Li, J. Cheng, Y. Yu, M. Wang, "Cell image segmentation based on an improved watershed algorithm", Proceedings of the 8th International Congress on Image and Signal Processing, Shenyang, China, 14-16 October 2015.
- [19] M.-H. Yang, D. Roth, N. Ahuja, "A Tale of Two Classifiers: SNoW vs. SVM in Visual Recognition", Springer-Verlag, Berlin Heidelberg, 2002, pp. 685-699.
- [20] B. Scholkopf, C. Burges, A. Smola, "Advances in Kernel Methods: Support Vector Learning", MIT Press, Cambridge, MA, 1999.
- [22] P. S. Hiremath, J. R. Tegnoor, "Follicle Detection and Ovarian Classification in Digital Ultrasound Images of Ovaries", Advancements and Breakthroughs in Ultrasound Imaging, 2013.

ResViT: A Framework for Deepfake Videos Detection

Original Scientific Paper

Wasim Ahmad

Institute of Information Sciences,
Academia Sinica, Taiwan
Department of Computer Science,
National ChengChi University, Taiwan
was_last@iis.sinica.edu.tw

Imad Ali

Department of Computer Science,
University of Swat, KP, Pakistan
imadali@uswat.edu.pk

Sahibzada Adil Shahzad

Institute of Information Sciences,
Academia Sinica, Taiwan
Department of Computer Science,
National Chengchi University, Taiwan
adilshah275@iis.sinica.edu.tw

Ammarah Hashmi

Institute of Information Science,
Academia Sinica, Taiwan
Institute of Information Systems and Applications,
National Tsing Hua University, Taiwan
hashmiammarah0@gmail.com

Faisal Ghaffar

System Design Engineering Department,
University of Waterloo, Canada
faisal.ghaffar@uwaterloo.ca

Abstract – Deepfake makes it quite easy to synthesize videos or images using deep learning techniques, which leads to substantial danger and worry for most of the world's renowned people. Spreading false news or synthesizing one's video or image can harm people and their lack of trust on social and electronic media. To efficiently identify deepfake images, we propose ResViT, which uses the ResNet model for feature extraction, while the vision transformer is used for classification. The ResViT architecture uses the feature extractor to extract features from the images of the videos, which are used to classify the input as fake or real. Moreover, the ResViT architectures focus equally on data pre-processing, as it improves performance. We conducted extensive experiments on the five mostly used datasets. Our analysis revealed that ResViT performed better than the baseline and achieved the prediction accuracy of 80.48%, 87.23%, 75.62%, 78.45%, and 84.55% on Celeb-DF, Celeb-DFv2, FaceForensics++, FF-Deepfake Detection, and DFDC2 datasets, respectively.

Keywords: deepfake, detection, GAN, vision transformer

1. INTRODUCTION

Deepfake is creating and manipulating videos, audio, or images produced by deep learning methods and techniques that appear real [1]. Lip-sync [2], puppet-master [3], and face swap [4] are some of the techniques used for synthetic video, image, and speech generation. With such technological advancement, the creation of deepfake videos, audio, and images is rising [5, 6]. According to the report of DeepTrace [7], in September 2019, approximately fifteen thousand fake videos were found, which was about two times higher than the previous year. These included about 96% pornographic, while 99% were female celebrities whose

faces were mapped on porn stars. Such deepfake videos may target famous personalities to denigrate a person, resulting in devastating damages. For example, deepfake videos can be used to destabilize the reputation of a political candidate by making the candidate appear to say or do things that never actually occurred.

Researchers are developing robust algorithms for differentiating real videos from fake ones to prevent this hazardous threat to society. For example, [8] and [9] tried to discover discrepancies in eye blinking for deepfake detection. To simulate eye blinking, [8] proposed a model that can be used to spawn the appearance of a face from a portrait. The same problem was also addressed by [9],

recommending a model that produces speaking videos with heads using facial expressions like eyes blinking. Brockschmidt *et al.* [10] proposed a facial forgery detection model for the detection of various spoofing methods, which helps detect reliably those detection methods that are invisible. FakeCatcher [11] is a deepfake detection technique that uses biological signals representing internal synthesizers and image generators. A convolution neural network model is proposed in [12] to identify the inconsistencies created during the creation of deepfakes.

However, these existing models for deepfake detections mainly focus on their architectures and ignore the importance of data pre-processing, which may improve the model performance [13]. Thus, training a model with proper data pre-processing techniques for detecting deepfake videos with higher accuracy. Moreover, these techniques lack generalizability for detecting deepfake. However, deep neural networks (DNNs) have shown superior performance in image classification compared to shallow layers [14, 15]; thus, carefully training a DNN model can get maximal deepfake artifacts for detecting deepfake videos with higher accuracy.

In this article, we propose, **ResViT**, which combines the **ResNet** model with the **Vision Transformer** to identify deepfake videos efficiently. The ResViT has generalized architecture as it extracts all local and global features of videos' frames (images) via ResNet and classifies it as fake or real via the attention mechanism of the vision transformer. Also, the ResViT architectures focus equally on data pre-processing, as it improves performance. Moreover, we train the proposed ResViT model on a diverse set of face images using the largest dataset currently available to detect deepfakes created in different settings, environments, and orientations. To evaluate ResViT, we conducted extensive experiments on the five mostly used datasets of Celeb-DF, Celeb-DFv2, FaceForensics++, FF-Deepfake Detection, and DFDC2. Our analysis revealed that ResViT performed better than the baseline and achieved the prediction accuracy of 80.48%, 87.23%, 75.62%, 78.45%, and 84.55% on Celeb-DF, Celeb-DFv2, FaceForensics++, FF-Deepfake Detection, and DFDC2 datasets, respectively. The main contributions of this article are as follows:

1. We propose the ResViT framework, which combines the ResNet model with the vision transformer to identify deepfake images efficiently.
2. We propose not only to focus on the architectures of ResViT but also on data pre-processing, as it improves performance.
3. We propose we train the ResViT model on a diverse set of face images using the largest dataset currently available to detect deepfakes created in different settings, environments, and orientations.

The rest of the article is organized as follows: Section 2 presents the literature review, while Section 3 presents the proposed framework. The experiments and results are described in Section 4. Section 5 concludes this article.

2. LITERATURE REVIEW

Deepfake can be created by switching two different identities in the visual stream, i.e., image or video (sequence of images). FakeApp [16] is the first deepfake technique that uses two autoencoders (AE) networks. An AE is an encoder-decoder architecture, Feedforward Neural Network (FFNN), that is trained self-supervised to reconstruct the input stream. The encoder downsamples the input in FaceApp and converts it to a latent representation called latent face features. The decoder mirrors the encoder and works reverse to upsample the latent representation to reconstruct the face images [17]. Face synthesis and face swapping are techniques used for fake videos. Using the face synthesis technique, it's possible to create unseen realistic images from training examples [18]. Application of Image synthesis and, more specifically, face image synthesis are face frontalization, face aging, and pose-guided generation.

Face synthesis can be done by generative adversarial networks (GANs), where we create a generative model responsible for creating a realistic face image. GAN-based architectures, e.g., StyleGAN [19], produce more realistic images that resemble the original images. There is a technique called FaceSwapping, which is a generative adversarial network-based method to generate deepfake videos. Face swap is the process of swapping or inserting the facial identity of the source image into the target image. This fake generation is used to insert actors in different video clips [12]. Traditional computer vision techniques and GANs based approaches synthesize face swaps. FSGAN and RSGAN are also used to perform face-swapping tasks.

Similarly, Face expressions can also be exchanged among individuals. The Face2Face technique manipulates facial expressions and projects source images onto some target faces in almost real-time without delay [20]—Face2Face synthesis images under different lighting and environmental conditions. The deep learning techniques for deepfake video detection have three main categories [21]. The first set of methods focuses on the psychological and physical behavior of the videos. It includes head pose movement and tracking eye blinking. The second type focuses on GANs' fingerprints and biological signals. The last category is data-driven and focuses on visual artifacts. [22] also proposed a CNN model that leverages the image transformation or augmentation (i.e., rotation, scaling, and shearing). A novel approach based on deep learning to detect forged videos was proposed in [23]. They mainly focus on facial reenactments, face swapping, replay attacks, and computer-synthesized image spoofing.

Transformers architecture is usually used for language processing-related tasks, and an immense number of its applications are available in the literature on natural language processing tasks. On the contrary, its application remains limited in image processing, video processing, and computer vision research. [24], shows that CNN-based architecture can be replaced by its alternative,

so-called transformers, which perform better. It outperformed as compared to state-of-the-art CNNs while requiring fewer computational resources for model training. Deepfake is synthesized by autoencoder (encoder + decoder) and generative adversarial models [25].

Unlike these works, we utilized ResNet, a more general model, for feature extraction and integrated it with the vision transformer. We also focus on data processing, making it easy for the vision transformer in classification.

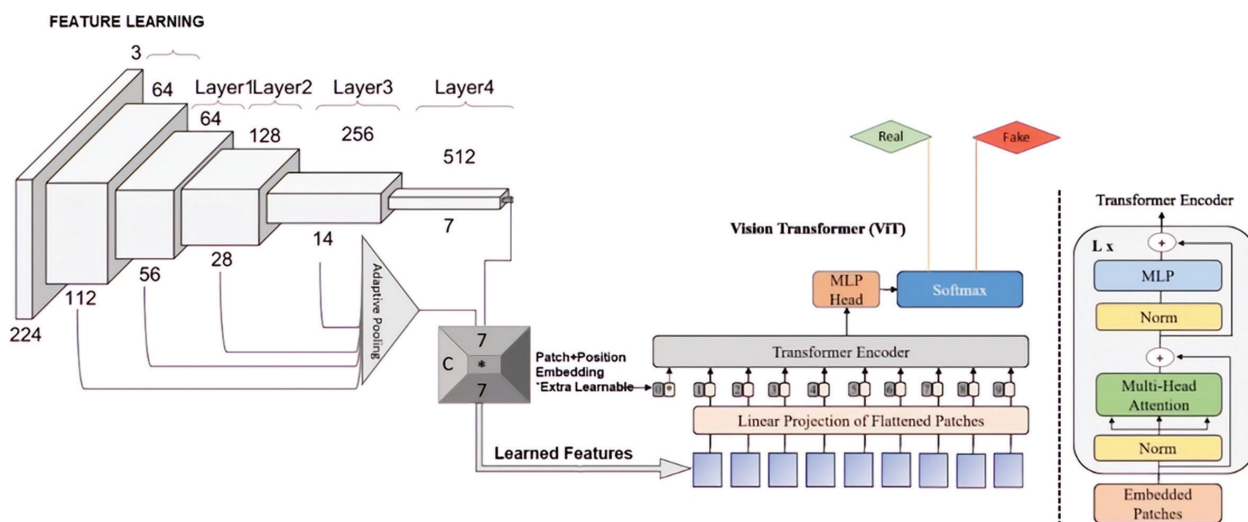


Fig. 1. The ResViT architecture.

Dataset	Generate Method	Total Videos	Actors	Train Images	Validation Images	Test Images
Celeb-DF	Deepfake	1203	13	10592	1245	1245
Celeb-DF-V2	Deepfake	6229	59	65936	7756	7756
FaceForensics++	Deepfake	2000	977	32499	3823	3823
FF-DFDC	Deepfake	1803	977	39727	4673	4673
DFDC2	Deepfake	2656	28	21088	2840	2840
Combined (FF)	Deepfake	2803	-	139879	16809	16809

Table 1. Datasets for deepfake detection

3. ResViT ARCHITECTURE

Fig. 1 shows the architecture of ResViT. The ResViT has feature learning and classification components. In the proposed ResViT, we use the ResNet 18 model to extract features from the images. The ResViT architecture works in a two-stage mechanism. Firstly, ResNet is used for extracting the features from the images. As shown in Fig. 1, we use the ResNet model to extract features from the images. We modify the ResNet model intermediate layers to get better image features. Each layer's output is the input of the next layer. We did not use the model's average pooling layer, flattening, and fully connected layers. Since we only need to extract the features, we do not use a fully connected layer. Finally, we return all four outputs and concatenate them. We reduce the dimensions to (Channels*7*7) and concatenate them. We apply different combinations of concatenating the outputs of the layer. Still, in some cases, we get a better model performance by concatenating the output of the first and last layer (x1 and x4) and the dimensions for that (3072*7*7).

Secondly, we used the vision transformer [24] for classification. Most natural language processing tasks

use transformers, primarily for sequential tasks. After better performance on many tasks, the transformer is also thought to be used for computer vision tasks. The vision transformers follow the mechanism of the earliest transformer with some minor input signal adjustment. These are the main components of our model ResViT. After we get the features from the ResNet, we map all those features to the transformer. Transformers take the input image in patches. Therefore, we need to divide our image into patches. We split the feature map into seven patches that are not fixed, and one can use any patch size. The patches are then entrenched into a linear sequence with the dimension of 1*1024. We need to perform position embedding so that each patch can be placed after the other. Therefore, we need to divide our image into patches. We split the feature map into seven patches that are not fixed, and one can use any patch size. The patches are then entrenched into a linear sequence with the dimension of 1*1024. We need to perform position embedding so that each patch can be placed after the other. Therefore, the patches are further added up into position embedding.

The dimension for position embedding then becomes 2*1024 in this case. Compared to the original

transformer, the vision transformer uses only the encoder. So, the patch and position embeddings are forwarded to the vision transformer. The transformer encoder has two blocks: Multiheaded Self-Attention (MSA) and Multi-Layer Perceptron (MLP), whose head and the job are similar to conventional CNN, as shown in Fig. 1. The input dimension has 2048 channels while the out channels are two, which signifies the two classes (fake and real). It has almost 40 million learnable parameters, and to get the final output, the MLP head has applied SoftMax, which alleviates the weight values between 0 and 1. It consists of a Feed Forward Network and is followed by a norm layer to normalize the interior layer. There are eight heads in the transformer. ReLU nonlinearity and a couple more layers are part of MLP.

4. EXPERIMENTS

In this section, we present the experimental setup to implement the model. We deliver the results achieved by our model implementation and interpret the experimental results.

4.1. DATASET

Deep learning models learn from data; thus, the dataset must be carefully prepared for higher prediction accuracy. Therefore, we pre-process the data so that the model learns all the features correctly. We use a couple of libraries like BlazeFace and MTCNN for face extraction, known as the rapid processing of large amounts of images. The dimensions and format of the extracted images are 224*224 and JPEG, respectively. Some examples of face extraction from real and fake datasets can be seen in Fig. 2.

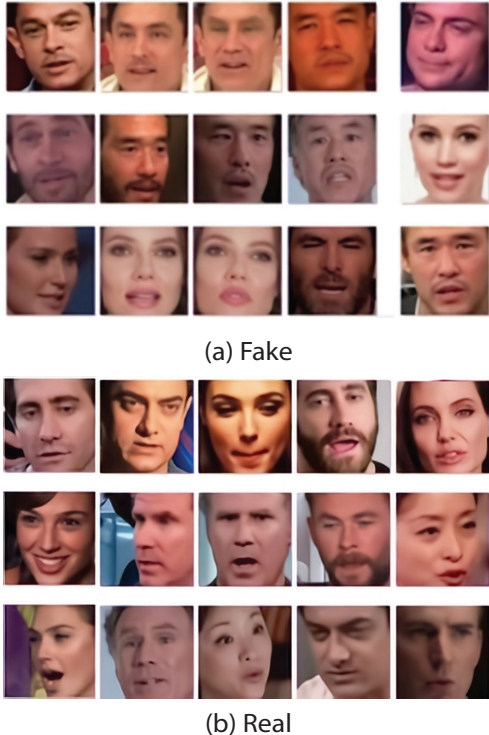


Fig. 2. Extracted frames from videos.

We split the datasets into training, validation, and testing. The data is then augmented using the library called Albumentation, known for transforming a huge amount of data. We train our model first to detect whether the video is real or fake. To make our model perform better, we train our model with large data. We tried to split the dataset so that maximum data could be provided for training the model. It took a lot of time to train the model and finetune it for better performance. After training the model, the validation data is provided, which is unseen data, and the model is finetuned repeatedly. Finally, the testing data is provided to the model to evaluate and predict whether our model classifies the video in relative class(fake/real) or not. We used Celeb-DF [2], Celeb-DFv2 [3], Faceforensics++ [4] and Deepfake Detection Datasets to evaluate our model. We combine the FaceForensic Deepfake and Deepfake Detection datasets. The number of training, validation, and testing images for each dataset is shown in Table 1.

4.2. EVALUATION

Before feeding our dataset to ResViT, we normalized and augmented the dataset at each iteration of the training phase. We use the learning rate of 0.001 and weight decay of 0.000001 for ten epochs. Once the model is trained, 30 images are forwarded to the model for the classification process. We calculate the accuracy of our model by using the log loss function. We used a binary cross-entropy function for calculating the loss. The purpose of the log loss function is to calculate the probability distribution between 0 and 1. The real class is represented with the value of $0 > y < 0.5$, while the fake class is represented with the value of $0.5 \geq y < 1$. For a fair comparison with the baseline, CViT [26], we trained our model with a batch size of 8 and 10 epochs under the same settings. The baseline model uses the VGG-16 architecture as a feature extractor and transformer as a classifier, using the DFDC dataset released by Facebook. We have limited resources; therefore, we did not use the DFDC dataset, which is approximately 470 Gigabytes.

We demonstrate our results by calculating the accuracies and losses for all datasets. We trained the baseline and the proposed models on all the datasets mentioned above and compared their results.

Fig. 3 shows the performance comparison of the ResViT on different datasets. The proposed ResViT performs better on each dataset and has achieved the prediction accuracy of 80.48%, 87.23%, 75.62%, 78.45%, and 84.55% on Celeb-DF, Celeb-DFv2, FaceForensics++, FF-Deepfake Detection, and DFDC2 datasets, respectively. The overall prediction accuracy of the ResViT on the combined datasets is 74.54.

Fig. 4 shows the performance comparison of the baseline and proposed ResViT under the same circumstances and resources. The figure shows that CViT has

achieved the prediction accuracy of 71.04%, 84.50%, 71.67%, 73.31%, and 73.42%, on Celeb-DF, Celeb-DFv2, FaceForensics++, FF-Deepfake Detection, and DFDC2 datasets, respectively, while the ResViT achieved higher accuracies of 80.48%, 87.23%, 75.62%, 78.45%, and 84.55%, on the same datasets. The overall prediction accuracy of the CViT on the combined datasets is 68.37%, while the ResViT has 74.54% prediction accuracy. The ResViT prediction accuracy is 15.79% higher than the CViT on the combined dataset. This is because our proposed ResViT architecture utilizes the ResNet model for feature extraction, which has better generalization than the CNN model. Moreover, the proposed

ResViT architecture focuses on pre-processing, which results in higher predictions than the baseline. Since we used the same settings as CViT, thus when ResViT results are better than the CViT, it automatically performs better than other baselines of the CViT.

Moreover, we present the training and validation losses and accuracy on three sample datasets for the proposed ResViT in Fig. 5 for the different number of epochs. Although we observe that the results improve with more epochs, we stick to the original settings of epochs of the CViT. We observe that the ResViT achieved better performance on all three sample datasets, as shown in Fig. 5 (a-e).

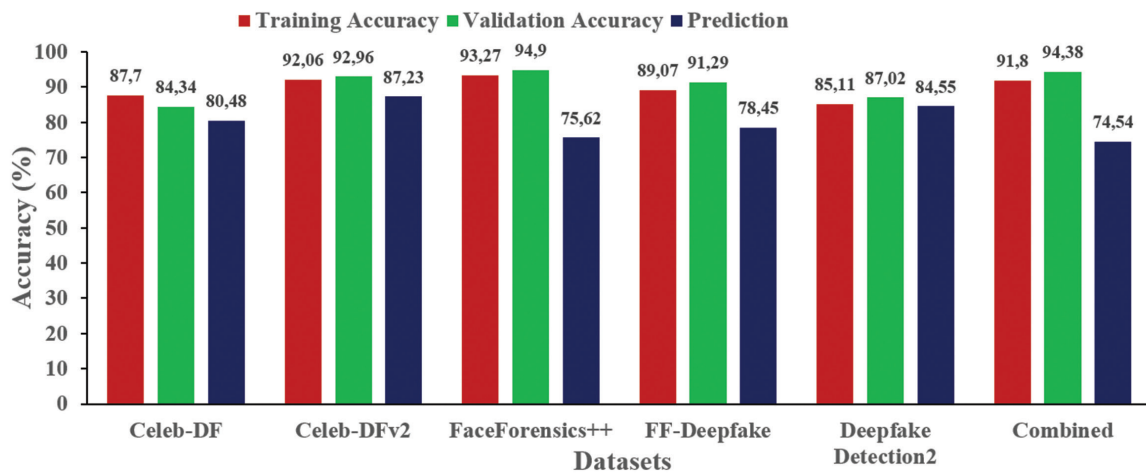


Fig. 3. ResViT performance on all datasets

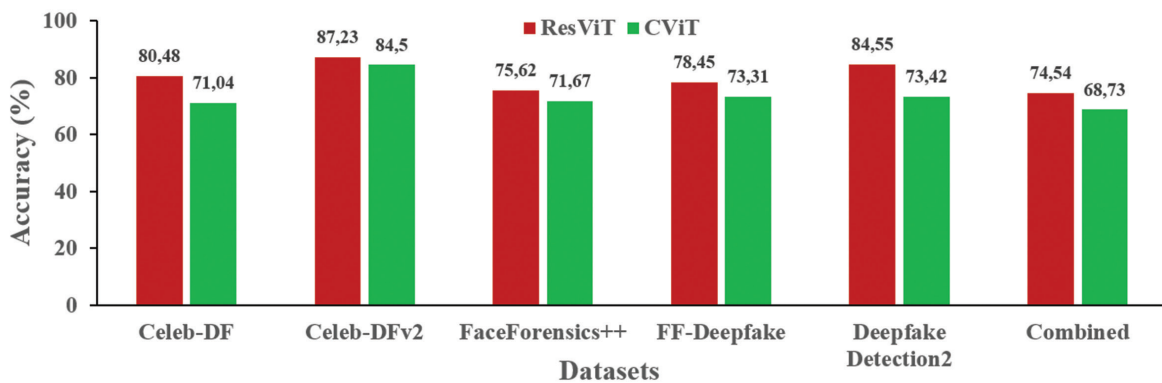
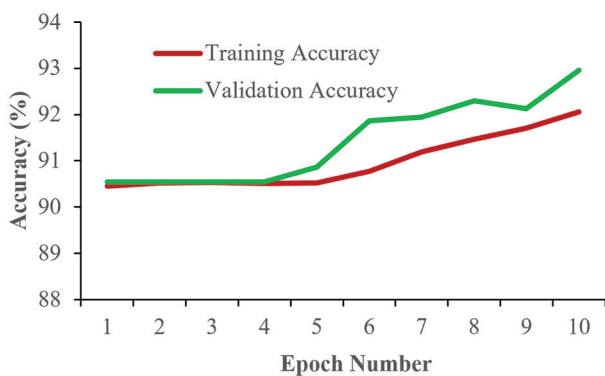
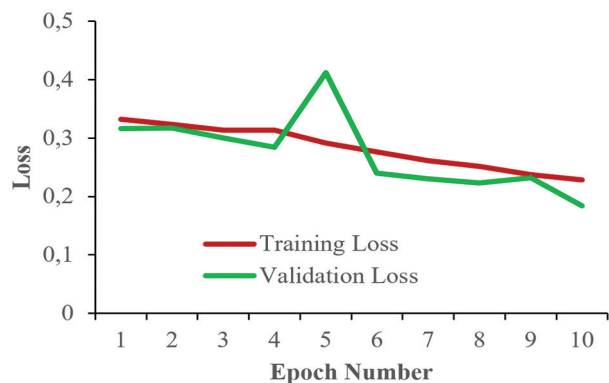


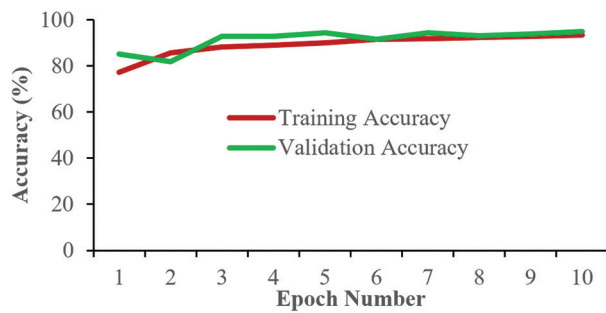
Fig. 4. ResViT and CViT prediction performance on all datasets.



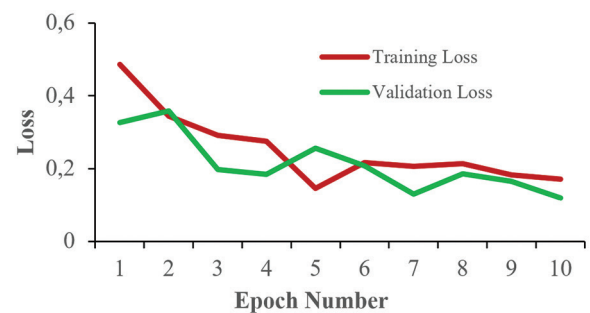
(a) Celeb-DFv2 training and validation accuracy



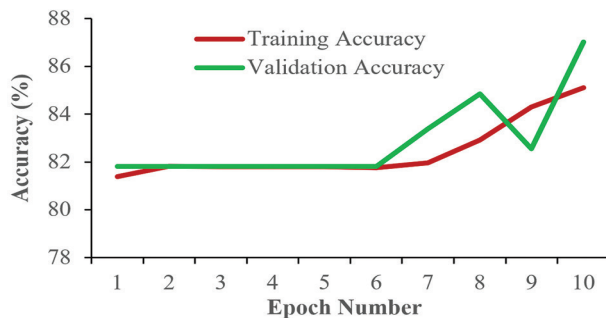
(b) Celeb-DFv2 training and validation loss



(c) FaceForensics++ training and validation accuracy



(d) FaceForensics++ training and validation loss



(e) DFDC2 training and validation accuracy



(f) DFDC2 training and validation loss

Fig. 5. ResViT validation losses and accuracy on three sample datasets for the different number of epochs.

5. CONCLUSION

In this article, we proposed ResViT, which combines the ResNet model with the Vision Transformer to identify deepfake videos efficiently. ResViT extracts all local and global features of videos via ResNet and classifies them as fake or real via the attention mechanism of the vision transformer. ResViT not only focuses on its architecture but also on pre-processing, which adds to higher prediction performance. We evaluated ResViT and baseline in the same settings with extensive experiments on the five mainly used datasets in deepfakes. We find that the proposed ResViT performs better than the baseline. We anticipated the better performance of ResViT, as the ResNet model has better generalization in feature extraction, and the pre-processing adds to prediction performance. Thus, such technology should be used to protect people, especially celebrities and politicians. In the future, we are determined to check the performance of the ResViT under massive datasets with more baseline models.

6. REFERENCES

- [1] T. Park, M.-Y. Liu, T.-C. Wang, J.-Y. Zhu, "Semantic Image Synthesis With Spatially-Adaptive Normalization", Proceedings of the IEEE/CVF Conference On Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15-20 June 2019, pp. 2337-2346.
- [2] P. KR, R. Mukhopadhyay, J. Philip, A. Jha, V. Namboodiri, C. Jawahar, "Towards Automatic Face-To-Face Translation", Proceedings of the 27th ACM International Conference on Multimedia, Nice France, 21-25 October 2019, pp. 1428-1436.
- [3] S. Suwajanakorn, S. M. Seitz, I. Kemelmacher-Shlizerman, "Synthesizing Obama: Learning Lip Sync From Audio", ACM Transactions on Graphics, Vol. 36, No. 4, 2017, pp. 1-13.
- [4] Y. Nirkin, Y. Keller, T. Hassner, "FSGAN: Subject Agnostic Face Swapping and Reenactment", Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 27 October - 2 November 2019, pp. 7184-7193.
- [5] B. Chesney and D. Citron, "Deep Fakes: A Looming Challenge For Privacy, Democracy, And National Security", The California Law Review, Vol. 107, 2019, pp. 1753.
- [6] E. Hsiang, "Deepfake: An Emerging New Media Object in the Age of Online Content", Boston University School of Law, 2020, Master Thesis.
- [7] L. Zheng, Y. Zhang, V. L. Thing, "A Survey on Image Tampering and Its Detection in Real-World Photos", Journal of Visual Communication and Image Representation, Vol. 58, No. 1, 2019, pp. 380-399.

- [8] H. X. Pham, Y. Wang, V. Pavlovic, "Generative Adversarial Talking Head: Bringing Portraits to Life With a Weakly Supervised Neural Network", arXiv:1803.07716, 2018.
- [9] K. Vougioukas, S. Petridis, M. Pantic, "Realistic Speech-Driven Facial Animation with GANs", *International Journal of Computer Vision*, Vol. 128, No. 5, 2020, pp. 1398-1413.
- [10] J. Brockschmidt, J. Shang, J. Wu, "On the Generality of Facial Forgery Detection", *Proceedings of the 16th IEEE International Conference on Mobile Ad Hoc and Sensor Systems Workshops*, Monterey, CA, USA, 4-7 November 2019, pp. 43-47.
- [11] U. A. Ciftci, I. Demir, L. Yin, "FakeCatcher: Detection Of Synthetic Portrait Videos Using Biological Signals", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020
- [12] Y. Li and S. Lyu, "Exposing Deepfake Videos By Detecting Face Warping Artifacts", arXiv:1811.00656, 2018.
- [13] P. Charitidis, G. Kordopatis-Zilos, S. Papadopoulos, I. Kompatsiaris, "Investigating The Impact of Pre-processing And Prediction Aggregation on the Deepfake Detection Task", arXiv:2006.07084, 2020.
- [14] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning For Image Recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27-30 June 2016, pp. 770-778.
- [15] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet Classification With Deep Convolutional Neural Networks", *Advances In Neural Information Processing Systems*, Tahoe, NV, USA, 3-6 December 2012, pp. 1-9.
- [16] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, ... C. M. Nguyen, "Deep Learning For Deepfakes Creation And Detection: A Survey", *Computer Vision and Image Understanding*, Vol. 223, 2022, pp. 103525,
- [17] M. A. Wani, F. A. Bhat, S. Afzal, A. I. Khan, "Advances in Deep Learning", First Edition, Springer Publisher, 2020.
- [18] H. Huang, P. S. Yu, C. Wang, "An Introduction to Image Synthesis With Generative Adversarial Nets", arXiv:1803.04469, 2018.
- [19] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, H. Li, "Protecting World Leaders Against Deep Fakes", *Proceedings of CVPR Workshops*, Long Beach, CA, USA, 15-21 June 2019, p. 38.
- [20] Y. Mirsky and W. Lee, "The Creation and Detection Of Deepfakes: A Survey", *ACM Computing Surveys*, Vol. 54, No. 1, 2021, pp. 1-41.
- [21] D. Afchar, V. Nozick, J. Yamagishi, I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network", *Proceedings of the IEEE International Workshop on Information Forensics and Security*, Hong Kong, China, 11-13 December 2018, pp. 1-7.
- [22] H. H. Nguyen, J. Yamagishi, I. Echizen, "Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, 12-17 May 2019, pp. 2307-2311.
- [23] D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, F. Zhu, "Deepfakes Detection with Automatic Face Weighting", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Virtual, 14-19 June 2020. pp. 668-669.
- [24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, S. Gelly, "An Image is Worth 16x16 Words: Transformers For Image Recognition at Scale", arXiv:2010.11929, 2020.
- [25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Y. Bengio, "Generative Adversarial Nets", *Advances in Neural Information Processing Systems*, Montreal, Canada, 8-11 December 2014, pp. 1-9.
- [26] D. Wodajo, S. Atnafu, "Deepfake Video Detection Using Convolutional Vision Transformer", arXiv:2102.11126, 2021.

Design of Super Twisting Integral Sliding Mode Control for Industrial Robot Manipulator

Original Scientific Paper

Shankar J Gambhire

Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur, AP 522302, India.
Faculty with ECE, MIT School of Engineering &
Sciences, MIT ADT University,
Pune(MH), India.
sjgambhire@gmail.com

D Ravi Kishore

Godavari institute of engineering and technology,
EEE, Rajahmundry, AP 533296. India.
dravikishore@gmail.com

Malligunta Kiran Kumar

Koneru Lakshmaiah Education Foundation, EEE,
Vaddeswaram, Guntur, AP 522302, India.
mkkumar@kluniversity.in

Sushant N Pawar

Ramrao Adik Institute of Technoloy, Navi Mumbai,
Maharashtra, India.
sushantnpawar@gmail.com

Abstract – In the present work, integral sliding mode based continuous control algorithm is extended to multi input multi output system. The typical integral sliding mode control (ISMC) contains nominal control with discontinuous feedback control due to which overall control becomes discontinuous in nature. The proposed controller is a fusion of two continuous terms and one of which is able to handle, estimate and reject the disturbance successfully. A proposed robust ISMC technique is applied for industrial robot manipulators which utilizes interactive manipulation activity. Here, robust position tracking control obtained via ISMC principle for two link IRM scheme influenced by parametric uncertainties and external disturbances. The proposed ISMC design replaces the discontinuous part by continuous control, which super twisting control is able to handle the disturbance rejection completely. The effectiveness of the proposed control technique is tested under uncertain conditions and comparison study with other controllers has been done. The simulation result shows that the tracking error is effectively minimized by the proposed technique in presence of uncertain conditions.

Keywords: Sliding mode control (SMC), Non-singularity, Super twisting control (STM), Industrial robotic manipulator (IRM)

1. INTRODUCTION

In recent years, industrial robotic manipulator (IRM) applications used to carry out advanced tasks in industrial automation like robotic surgery, welding, painting, polishing, laser cutting work etc. have added great importance. Accurate besides precise robot manipulators motion control is required to perform such classy tasks. Hence, design of motion control schemes for IRM systems has emerged as an active research area. Because of its complex dynamics comprising time changing dynamical-structure in addition strong dynamic coupling, inherent non-linearities, parameter uncertainties a dedicated motion control structure designing for IRM scheme is tough task [1]. Several advanced control schemes studied comprehensively by a number of researchers over a period of time. These are mostly comprises continuous sliding mode control, feedback linearization [2], decentralized control, model predictive control, adaptive control, fuzzy and neural network

control. Intelligent control method like fuzzy plus neural network is good option for addressing nonlinear plus uncertain IRM dynamics. However, it embroil many of design parameters and complex rules that would lead to complex design procedure and sometimes impossible for implementation point of view.

Among all said methods, sliding mode control (SMC) is popular, robust control and widely executed for many input many output (MIMO) linear plus nonlinear systems. Robustness property in presence of external disturbances and parameter uncertainties is the primary feature of SMC. A variety of second order sliding mode control algorithms are discussed in [3-5, 15-19], one of the most potential algorithms of all of them is super twisting. This approach is developed to avoid chattering in systems having relative degree of one. The twisting around the origin characterizes trajectory on the two sliding planes. Super twisting control is a continuous control that ensures all key features of first order SMC for

system with smooth matched bounded uncertainties or disturbances. Super twisting algorithm with initial stability findings are majorant curve based concept. Thereafter, the Lyapunov scheme was introduced for proofs by [6]. From the several SMC approaches, integral SMC [7-14,17], may be readily integrated other new latest control strategies like proportional integral derivative (PID) control, linear feedback control, model predictive control, optimum control and so on, while retaining their features. As a result, ISMC may be described as bridge in between other control methods plus sliding mode which has greater robustness than other existing control systems. Though ISMC only has one issue, it has a discontinuous portion of control. This issue will be resolved if control is held continuously.

In this work, ISMC has 2 parts viz. first nominal control $u_{nominal}$ next continuous control $u_{continuous}$ or discontinuous control (given by Utkin). To obtain desired trajectory for a system nominal-control is designed (without disturbance). $u_{nominal}$ and $u_{continuous}$ design are totally sovereign. It may be noted that once trajectories on the sliding surface, continuous control acts like a disturbance observer with its value equal to negative of the disturbance. Thus $u = u_{nominal} + u_{continuous}$ is applied to structure with disturbance, $u_{continuous}$ discards disturbance plus desired trajectory is obtained through application of $u_{nominal}$.

This work considers a continuous SMC algorithm based on ISMC. Continuous control is replaced with ISMC discontinuous part. The property of disturbance rejection of super twisting control is utilized here by replacing discontinuous control with super twisting control. Discontinuous term will result in chattering, in addition it is undesirable from practical execution point of view. Presented algorithm has been implemented for two link IRM (Fig.1.).

Here robust position tracking control pattern is planned in addition realistic for 2 link IRM method using condition which is uncertain. Significant contribution is given as below

- Convergence speed is accelerated as well as chattering is reduced by replacing with a continuous control with discontinuous part of ISMC.
- The property of disturbance rejection of super twisting control is utilized here by replacing discontinuous control with super twisting control.
- There are two phases in sliding mode. I. reaching phase, where system states are driven to the switching manifolds from any initial state in finite time and II. Sliding phase, where system is induced into sliding-motion on switching manifold, i.e. it becomes attractor.
- Singularity free control is possible with the proposed sliding manifold.
- An effective disturbance estimator as a super twisting control is presented to counteract unexpected unmodeled dynamics impacts, unidentified parameters owing to uncertainties in the

parameters plus measurement noises of the IRM system.

- In this study, it is not necessary to have former knowledge of upper bounds of lumped disturbance.
- This control provides greater robustness against lumped uncertainty while maintaining excellent tracking precision.

This technique is tested to verify using a generic 2 link IRM model meant for monitoring predefined complex trajectory in indefinite work environment.

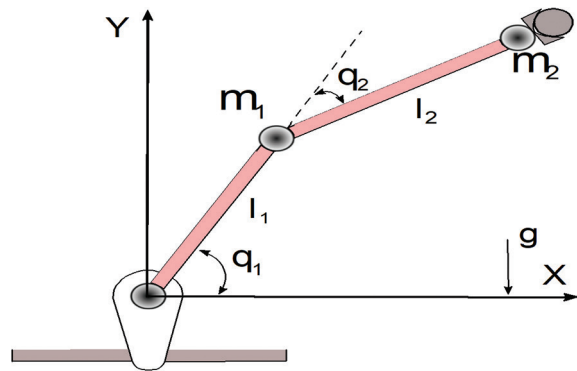


Fig.1. Two link industrial robotic manipulator

2. PROBLEM DEFINITION

For n-link IRM system equation for dynamic motion may be given as

$$G(q) + C(q, \dot{q})\dot{q} + M(q)\ddot{q} = \tau_{ex} + \tau \quad (1)$$

Given: $\dot{q}, q, \ddot{q} \in \mathbb{R}^{n \times 1}$ vectors of velocity, position plus acceleration of joints of IRM separately. Applied joint torques vector: $\tau \in \mathbb{R}^{n \times 1}$, $\tau_{ex} \in \mathbb{R}^{n \times 1}$ unknown external disturbances vector, $M(q) \in \mathbb{R}^{n \times n}$ IRM inertia matrix, $C(q, \dot{q})\dot{q} \in \mathbb{R}^{n \times n}$ centripetal and Coriolis forces matrix, gravitational forces vector: $G(q) \in \mathbb{R}^{n \times 1}$.

Uncertainties of IRM model matrices, i.e.

$$\begin{aligned} \Delta M(q) &= M(q) - M_0(q) \\ \Delta C(q, \dot{q}) &= C(q, \dot{q}) - C_0(q, \dot{q}) \\ \Delta G(q) &= G(q) - G_0(q) \end{aligned} \quad (2)$$

Given: $M_0(q)$, $C_0(q, \dot{q})$ and $G_0(q)$ are the nominal terms; $\Delta M(q)$, $\Delta C(q, \dot{q})$ and $\Delta G(q)$ model matrices perturbations.

So IRM dynamic model takes form as

$$M_0(q)\ddot{q} + C_0(q, \dot{q})\dot{q} + G_0(q) = \tau + \tau_d \quad (3)$$

A term (Lumped-uncertainty vector) given $\tau_d = \tau_{ex} + H(q, \dot{q}, \ddot{q})$ besides $H(q, \dot{q}, \ddot{q}) = -\Delta M(q)\ddot{q} - \Delta C(q, \dot{q})\dot{q} - \Delta G(q)$ which is circumscribed thru subsequent function

$$\|H(q, \dot{q}, \ddot{q})\| \leq p_0 + p_1 \|q\| + p_2 \|\dot{q}\|^2 \quad (4)$$

where positive constants p_2, p_1, p_0

Eq. 1 known as dynamic equation possesses properties useful in controller design plus to analyses stability

(P: 1) $M(q)$ is positive beside symmetric plus bounded by $m_1 \leq \|M(q)\| \leq m_2$ where constants $m_1, m_2 \geq 0$

(P: 2) Skew symmetric matrix is $\dot{M}(q) - 2C(q, \dot{q})$

(P: 3) Bounded matrix $C(q, \dot{q})$ gives $\|C(q, \dot{q})\| \leq \lambda_0 \|\dot{q}\|$ for $\lambda_0 > 0$

3. CONTROLLER DESIGN

3.1. SUPER TWISTING MULTIVARIABLE (STM) ALGORITHM

Super twisting algorithm is designed to regulate the schemes in presence of lumped uncertainty and as well it reduces the chattering phenomenon. It gives continuous control signal that assures all of fundamental features of first-order SMC, given first derivative of the matched disturbance is finite. Improvement on well-known super twisting technique is STM method. It may be utilized for MIMO system control without system states decoupling.

Let us take MIMO scheme as,

$$\dot{x} = B(d + u) \quad (5)$$

where B, x is invertible as well as known, u as a control and d is the bounded but unknown uncertainty.

The term d involves, $d = d_1 \|x\| + d_2$ where $d_1 \leq \Delta_1$ besides $d_2 \leq \Delta_2$

Super twisting multivariable control u is given below

$$u = B^{-1} \left[-k_1 \frac{x}{\|x\|^2} + v - k_2 x \right]$$

$$\dot{v} = -k_3 \frac{x}{\|x\|} - k_4 x \quad (6)$$

where $k_i > 0, i=1,2,3,4$ are the constant parameters chosen in way that the aforementioned controller would be able to stabilize the MIMO structure in finite-time. Putting u as of (6) to (5), we can write

$$\dot{x} = -k_1 \frac{x}{\|x\|^2} + v - k_2 x + d_1 \|x\| + d_2$$

$$\dot{v} = -k_3 \frac{x}{\|x\|} - k_4 x \quad (7)$$

By specifying $\eta = v + d_2$, we get (7)

$$\dot{x} = -k_1 \frac{x}{\|x\|^2} - k_2 x + d_1 \|x\| + \eta$$

$$\dot{\eta} = -k_3 \frac{x}{\|x\|} - k_4 x + \dot{d}_2 \quad (8)$$

State variables finite-time convergence x, \dot{x} as well η are to be achieved and keep zero for subsequent time by k_1, k_2, k_3, k_4 . As of eq (8) when states touches origin we get, $\eta = v + d_2 = 0$

$$d_2 = \eta \quad (9)$$

Because of the aforementioned characteristic, the multivariable super twisting algorithm is used as both

a controller and a disturbance estimator. In the next part, we will present continuous ISMC for MIMO systems based on the previously described disturbance observation characteristic.

3.2. ROBUST INTEGRAL SMC (ISMC) FOR ROBOT MANIPULATOR SCHEME

The ISMC consists of two parts: u nominal control ($u_{nominal}$), besides (II) continuous or discontinuous control ($u_{continuous}$). The nominal control is aimed at keeping the system on the planned trajectory in the absence of disturbance. The concepts of $u_{nominal}$ plus $u_{continuous}$ are entirely separate. It should be highlighted that once sliding-surface with trajectories ($= 0$ as per property of multi variable super twisting disturbance observation) are established, continuous control works as disturbance estimator, with its value equal to negative of disturbance. As a result, when $u = u_{nominal} + u_{continuous}$ is applied to a system with a disturbance, $u_{continuous}$ castoffs disturbance besides $u_{nominal}$ produces required trajectory.

This technique is mathematically described as follows,

Ponder the structure

$$\dot{x} = Ax + B(d + u) \quad (10)$$

Given A, B, u, x, d are system-matrix, input matrix, control, state plus disturbance separately. Depending on the system's nature, the input matrix in addition system matrix might be nonlinear or linear. The control input for the system (10) is to be considered as

$u = u_{nominal} + u_{continuous}$, where $u_{nominal}$ is acting as a servo control and $u_{continuous}$ is acting as a regulatory control under the effect of lumped disturbances. Here, $u_{nominal}$ can be chosen as PID, any multivariable linear control, state-feedback, linear quadratic regulator (LQR) optimal control, adaptive-control, time variant control, and so on. The systems (10) required sliding surface defined as

$$s = [x(t) - x(t_0) - \int_0^t (Ax + B u_{nominal}) d\tau] G \quad (11)$$

Where, projection-matrix plus initial-condition of scheme $G, x(t_0)$ the sliding surface is designed so that system trajectories begin from it, plus if a disturbance comes into play, $u_{continuous}$ becomes functional and disturbances are corrected. It would be explicated mathematically as

$$\dot{s} = G[Ax + Bu + d - Ax - Bu_{nominal}]$$

$$= G[Ax + B(u_{nominal} + u_{continuous} + d) - Bu_{nominal}]$$

$$= GB[u_{continuous} + d] \quad (12)$$

Assume devoid of loss of generality $GB = \Phi^{m \times m}$, given n dimensional unit square matrix I (or else $u_{continuous}$ is $((GB)^{-1})$ scaled control, thus

$$\dot{s} = [u_{continuous} + d] \quad (13)$$

Assuming $u_{continuous}$ is formed using STM control

$$u_{continuous} = -k_1 \frac{s}{\|s\|^{\frac{1}{2}}} + v - k_2 s$$

$$\dot{v} = -k_3 \frac{s}{\|s\|} - k_4 s \quad (14)$$

where $k_i > 0, i=1, 2, 3, 4$ are chosen appropriately by considering the stabilization of system. After substituting $u_{continuous}$ from (14) to (13), we can get

$$\dot{s} = -k_1 \frac{s}{\|s\|^{\frac{1}{2}}} + v - k_2 s + d$$

$$\dot{v} = -k_3 \frac{s}{\|s\|} - k_4 s \quad (15)$$

By specifying $\eta=v+d$ we can write (15)

$$\dot{s} = -k_1 \frac{s}{\|s\|^{\frac{1}{2}}} - k_2 s + \eta$$

$$\dot{v} = -k_3 \frac{s}{\|s\|} - k_4 s + d \quad (16)$$

Remark 1. s and v initial conditions essential for the solution of system eq. (15). We have previously designed s due to its zero initial conditions, since v is a fictional variable, it is always possible to pick $v = 0$ as an initial condition. Therefore, in order to initiate the 2nd order SMC from the beginning moment, starting values of s and \dot{s} need to be zero. The starting value of converted variable η contains initial value of disturbance d shown in (16). All the time it is difficult to estimate starting condition of the disturbance d , with the exception of some special circumstances where there is no disturbance at all, such as fault diagnosis difficulties when no fault there at all, which indicate $d = 0$. Hence, $\eta = 0$ for $t \geq 0$ from now both s plus \dot{s} are zero from the beginning moment. Therefore, when there is nonzero starting disturbance then sliding mode will begin subsequently some finite time $t \geq \tau$ as $z \neq 0$ this may be made very trivial if possible with selecting apposite values of $k_i = 1, 2, 3, 4$. Hence, when s and fictitious variable η are achieved to zero, so \dot{s} turn into zero and it relics zero all the time even in presence of disturbance.

As per eq. (16), we can make conclusive remark as $s = \eta = 0$ which implies $\dot{s} = 0$ in finite time for appropriate values of gains $k_i = 1, 2, 3, 4$. In the final step, from eq (16) $v = -d$. And from eq (14) $u_{continuous} = v = -d$. This indicates that while the scheme is in sliding mode, the value of the 'disturbance' $d = -v$ and is cancelled out. Therefore, when it is on a sliding surface, the system guided by nominal control then designed to be stable.

4. PERFORMANCE ANALYSIS PLUS NUMERICALL SIMULATIONS

With general two link IRM model performane of existing procedure has been appraised here [14] it gives

$$M(q) = \begin{bmatrix} a_{11}(q_2) & a_{12}(q_2) \\ a_{12}(q_2) & a_{22} \end{bmatrix},$$

$$C(q, \dot{q}) = \begin{bmatrix} -b_{12}(q_2)\dot{q}_1^2 - 2b_{12}(q_2)\dot{q}_1\dot{q}_2 \\ b_{12}(q_2)\dot{q}_2^2 \end{bmatrix},$$

$$G(q) = \begin{bmatrix} c_1(q_1, q_2)g \\ c_2(q_1, q_2)g \end{bmatrix}, \tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}, \tau_{ex} = \begin{bmatrix} \tau_{ex1} \\ \tau_{ex2} \end{bmatrix} \quad (17)$$

Where:

$$a_{11}(q_2) = (m_1 + m_2)l_1^2 + m_2l_2^2 + 2m_2l_1l_2 \cos(q_2) + J_1,$$

$$a_{12}(q_2) = m_2l_2^2 + m_2l_1l_2 \cos(q_2),$$

$$a_{22} = m_2l_2^2 + J_2,$$

$$b_{12}(q_2) = m_2r_1r_2 \sin(q_2),$$

$$c_1(q_1, q_2) = (m_1 + m_2)l_1 \cos(q_2) + m_2l_2 \cos(q_1 + q_2),$$

$$c_2(q_1, q_2) = m_2l_2 \cos(q_1 + q_2)$$

The IRM model parameters are shown in Table 1.

The high frequency effect of external disturbances plus measurement noises are considered in simulation for presenting the effectiveness of control in uncertain environment. The same external disturbances as in eq (18) used in simulation.

$$\tau_{ex1} = 2 \sin(t) + \frac{1}{2} (\sin(200\pi t))$$

$$\tau_{ex2} = \cos(2t) + \frac{1}{2} (\sin(200\pi t)) \quad (18)$$

Table 1. Model parameters of IRM

Symbol	Definition	Value
l_1	Length of link 1	1 m
l_2	Length of link 2	0.85 m
J_1	Moment of Inertia of Motor 1	5 kg m
J_2	Moment of Inertia of Motor 2	5 kg m
m_1	Mass of Link 1	0.5 kg
m_2	Mass of Link 2	1.5 kg
\hat{m}_1	Estimated Mass of Link 1	0.4 kg
\hat{m}_2	Estimated Mass of Link 2	1.2 kg
G	Gravitational Constant	9.8 m/s ²

Also the control behavior is tested in presence of parametric uncertainty by introducing 20 % parameter uncertainties in model matrices of the IRM.

IRM initial values elect as $q_1(0)=1.0, q_2(0)=1.5, \dot{q}_1(0)=0$ and $\dot{q}_2(0)=0$

IRM system chosen trajectory to be trailed is picked as

$$q_d = [q_{d1}, q_{d2}] \text{ with}$$

$$q_{d1} = 1.25 - \left(\frac{7}{5}\right) e^{-t} + \left(\frac{7}{20}\right) e^{-4t}$$

$$q_{d2} = 1.25 + e^{-t} - \left(\frac{1}{4}\right) e^{-4t} \quad (19)$$

Comparison of proposed method with benchmark controllers like conventional SMC, computed-torque controller (CTC) in addition to this a non singular terminal SMC are considered. The corresponding control laws are mentioned as below,

$$\tau = M_0(q) (K_p e(t) + K_v \dot{e}(t) + \ddot{q}_d) + C_0(q, \dot{q}) + G_0(q)$$

:CTC

$$\tau = M_0(q) (K_1 \text{sat}(s(t), \phi) + K_2 s(t) + \alpha \dot{e}(t) + \ddot{q}_d) + C_0(q, \dot{q}) + G_0(q)$$

:SMC

$$\tau = [M_0(q) (\ddot{q}_d - \beta^{-1} \gamma^{-1} \text{sig} \dot{e}^{2-\gamma} - K_1 \text{sat} - K_2 \text{sig}(s)^\rho) + C_0(q, \dot{q}) + G_0(q)]$$

:NTSMC

$$\tau = -\eta_1 \left(\frac{e}{\|e\|^{3/2}} \right) - \eta_2 \left(\frac{\dot{e}}{\|\dot{e}\|^{3/2}} \right) - k_1 \left(\frac{s}{\|s\|^{3/2}} \right) v - k_2 s$$

: Proposed

CTC: nominal controller settings: - $K_p = 25I_2$, $K_v = 10I_2$, SMC: $\alpha=6I_2$, $k_1 = 4.3I_2$, $k_2 = 2.6I_2$, $\phi=0.009I_2$, thru $(t)+\dot{e}(t)=s(t)$, NTSMC: $k_1=k_2=2.5$, $\beta=1I_2$ and $\gamma=1.5$ using $\beta \text{sig}(e(t))^\gamma + e(t) = s(t)$. Intended $\eta_1=35$, $\eta_2=15$, $k_1=2.1$, $k_2=40$, $k_3=2.5$ and $k_4=60$

With suggested robust ISMC numerical simulation results obtained for this effort as per in Eq. (14) with Figs 2-9. Fig 2 and 3 explains joint 1 and joint 2 position tracking control each. Offered control tactic not only show the tight control performance but also quickly follow the desired trajectory all through simulation run. NTSMC over SMC and CTC for position tracking control provides small-steady state error nevertheless NTSMC [14] gives better tracking results as related to SMC and CTC as shown in Figs 4-5. Even though projected proposed robust ISMC control provides all through the run zero steady state error in presence of parameter uncertainties plus time varying external disturbance. The corresponding control torque generated during position tracking by following controllers are shown in Figs 6-7. It should be noted that projected control requires minimum control torque among all controllers, to attain perfect position tracking control. Figs 8-9 clearly shows disturbance estimation in both the joints.

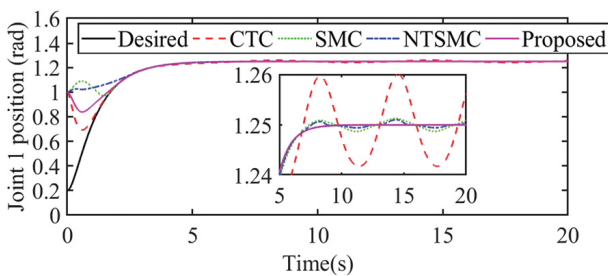


Fig. 2. Joint 1 Position Tracking

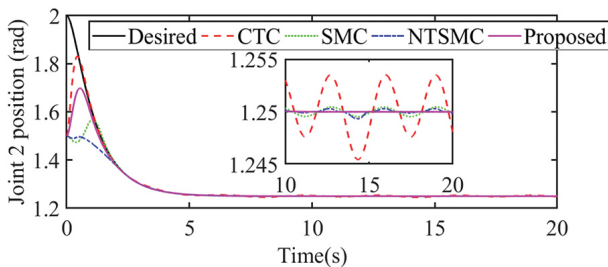


Fig.3. Joint 2 Position Tracking

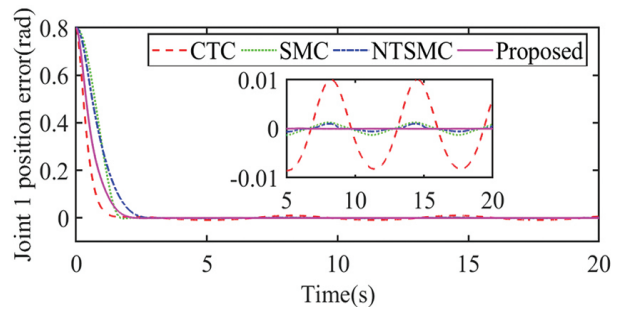


Fig. 4. Joint 1 Position Tracking Error

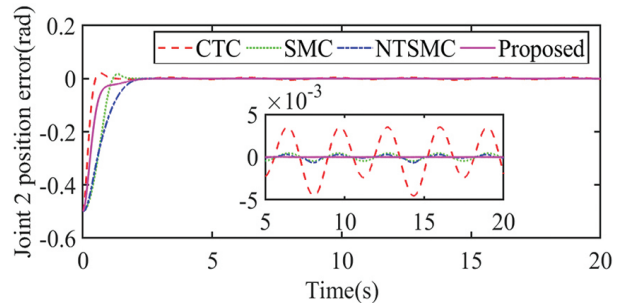


Fig. 5. Joint 2 Position Tracking Error

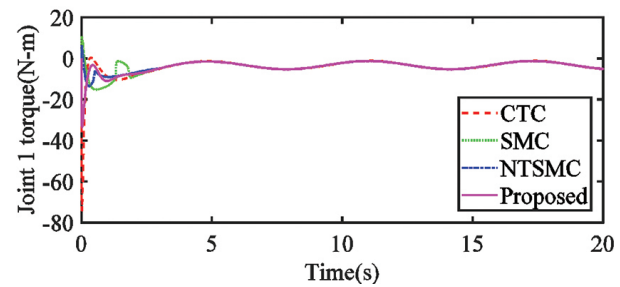


Fig. 6. Joint 1 Position Tracking Control Input

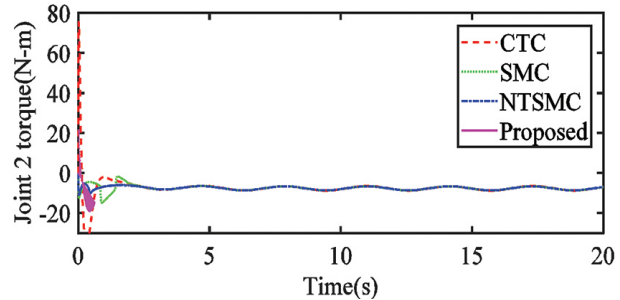


Fig. 7. Joint 2 Position Tracking Control Input

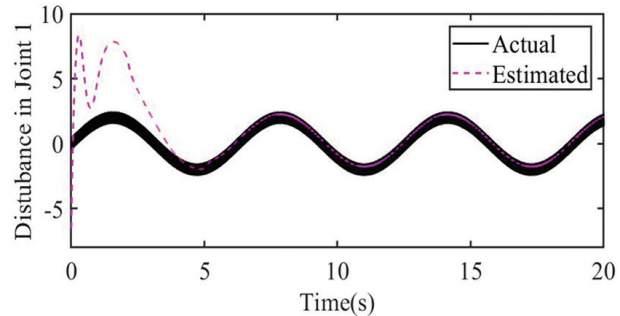


Fig. 8. Disturbance Estimation in Joint 1

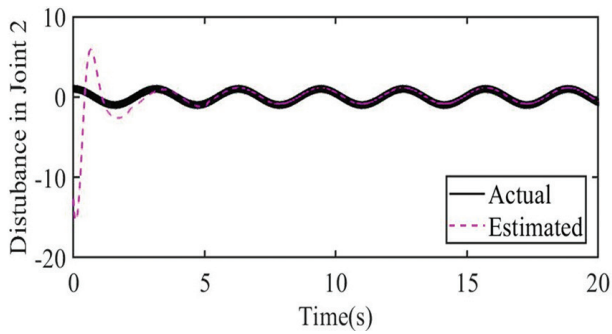


Fig. 9. Disturbance Estimation in Joint 2

Performance measures appraising quantitative performance of controllers castoff shown below;

- a) Root mean square (RMS) error

$$RMS = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}}$$

- b) Norm of the error (Euclidean) L^2

$$L^2 \text{ norm} = \sqrt{\sum_{i=1}^n e_i^2}$$

- c) Absolute error value Integral (IAE)

$$IAE = \int |e(t)| dt$$

- d) Using time interval integration then multiplication thru absolute error value (ITAE)

$$ITAE = \int t|e(t)| dt$$

Table 2 shows computed performance indexes .

As compared to other control schemes from these indices, it can be confirmed that projected controller has excellent dynamic performance, using agitation estimator in control regulation negates lumped uncertainty effects per sampling instant.

In overall, using disturbance estimator finite and fast convergence property thru enriched robustness is strength of this control structure .As per use of perturbation from dynamics of terminal sliding-manifold estimator estimates uncertainties together with unknown nonlinear dynamics of the manipulator in addition external disturbances. So control input compensate for uncertainty which exists throughout position trajectory tracing of IRM system, at every sampling period. In consequence it eliminates the need for information about the bounds of the existing lumped disturbance vector. Efforts displays merely one restriction that initial control inputs nevertheless that would be resolved through employing computer provided soft limiters, which are used to smoothen the signal known as low pass filters. It may be simply extended to any kind to robotic system together with serial as well parallel robots due to simplicity in design control procedure.

Table 2. IRM system position tracking control quantitative analysis with initial condition $q_1(0)=0.2$, $q_2(0)=2$

Control Laws	e1				e2			
	RMS	L^2 norm	IAE	ITAE	RMS	L^2 norm	IAE	ITAE
CTC	0.0923	41.2572	4.3737×10^3	1.2118×10^8	0.0441	22.2635	2.5851×10^3	4.5228×10^7
SMC	0.1430	63.9588	6.9782×10^3	4.9020×10^8	0.0784	35.0708	3.3289×10^3	1.8469×10^7
NTSMC	0.1376	61.5485	7.2406×10^3	5.0729×10^7	0.0799	35.7276	3.8895×10^3	2.3393×10^7
PROPOSED	0.0390	21.0201	2.2565×10^3	8.1910×10^6	0.0210	15.2050	0.9820×10^3	3.2210×10^6

5. CONCLUSION

This work has proven the effect of robust integral sliding mode control to the IRM system in presence of uncertain conditions. The replacement of discontinuous term with continuous super twisting control gives the multi feature of control as well as disturbance estimation and rejection. The proposed control scheme efficiency has been tested by comparing with well-known controllers. The proposed technique has several advantages, including higher accuracy, faster finite-time convergence, singularity free control plus no need for prior knowledge of the uncertainty bounds. The suggested solution also increases closed-loop system overall stability in presence of lumped disturbances due to super twisting control term. The design method presented here is based on a rational design technique that considers trade-off between response to set-point changes, stability of overall process with controller and

robustness to plant parameter variations. The present work can be extended in future directions by noting the features like, SMC generally effectively handles the matched disturbance part of the plant and observer based structure is able to handle the mismatched disturbance part of the system. Also this kind of structure of the control can be extended to other actuated and under actuated systems.

6. REFERENCES

- [1] Z. Li, Y. Kun, B. Stjepan, X. Bugong, "On motion optimization of robotic manipulators with strong nonlinear dynamic coupling using support area level set algorithm", International Journal of Control, Automation & Systems, Vol. 11, No. 6, 2013, pp. 1266-1275.

- [2] K. Kreutz, "On manipulator control by exact linearization", *IEEE Transactions on Automatic Control*, Vol. 34, No. 7, 1989, pp. 763-767.
- [3] L. Fridman, A. Levant, "Higher order sliding modes", *Sliding mode control in engineering*, Vol. 11, 2002, pp. 53-102.
- [4] A. Levant, "Homogeneity approach to high-order sliding-mode design", *Automatica*, Vol. 41, No. 5, 2005, pp. 823-830.
- [5] A. Levant, "Principles of 2-sliding mode design", *Automatica*, Vol. 43, No. 4, 2007, pp. 576-586.
- [6] J. A. Moreno, M. Osorio, "Strict Lyapunov functions for super twisting algorithm", *IEEE Transactions on Automatic Control*, Vol. 57, No. 4, 2012, pp. 1035-1040.
- [7] Z. Li, F. Wang, D. Ke, J. Li, W. Zhang, "Robust continuous model predictive speed & current control for pmsm with adaptive integral sliding mode approach", *IEEE Transactions on Power Electronics*, Vol. 36, No. 12, 2021, pp. 14398-14408.
- [8] X. Zhang, Y. Huang, Y. Rong, G. Li, H. Wang, C. Liu, "Recurrent neural network based optimal integral sliding mode tracking control for four wheel independently driven robots", *IET Control Theory & Applications*, Vol. 15, No. 10, 2021, pp. 1346-1363.
- [9] W. Ji, Q. Jianbin, H. R. Karimi, Y. Fu, "New results on fuzzy integral sliding mode control of nonlinear singularly perturbed systems", *IEEE Transactions on Fuzzy Systems*, Vol. 29, No. 7, 2020, pp. 2062-2067.
- [10] S. Yu, M. Xie, H. Wu, J. Ma, Y. Li, H. Gu, "Composite proportional-integral sliding mode control with feedforward control for cell puncture mechanism with piezoelectric actuation", *ISA transactions*, Vol. 124, 2020, pp. 427-435.
- [11] L. Zhao, L. Dai, Y. Xia, P. Li, "Attitude control for quadrotors subjected to wind disturbances via active disturbance rejection control & integral-sliding mode control", *Mechanical Systems & Signal-Processing*, Vol. 129, 2019, pp. 531-545.
- [12] S. Yu, X. Yu, B. Shirinzadeh, Z. Man, "Continuous finite time control for robotic manipulators with terminal sliding mode", *Automatica*, Vol. 41, No. 11, 2005, pp. 1957-1964.
- [13] D. Zhao, S. Li, F. Gao, "A new terminal sliding mode control for robotic manipulators", *International Journal of Control*, Vol. 82, No. 10, 2009, pp. 1804-1813.
- [14] J. Yuh, "Design and control of autonomous underwater robots", *A survey. Autonomous Robots*, Vol. 8, No. 1, 2000, pp. 7-24.
- [15] S. Kolahi, M. Reza, M. R. Gharib, A. Heydari, "Design of a non-singular fast terminal sliding mode control for second-order nonlinear systems with compound disturbance." *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, Vol. 235, No. 24, 2021, pp. 7343-7352.
- [16] S. J. Gambhire, D. R. Kishore, P. S. Londhe and S. N. Pawar, "Review of sliding mode based control techniques for control system applications." *International Journal of Dynamics and Control*, Vol. 9, 2021, pp. 363-378.
- [17] H. Xiao, D. Zhao, S. Gao and S. K. Spurgeon, "Sliding mode predictive control: A survey " *Annual Reviews in Control*, 2022. (in press)
- [18] J. Davila, L. Fridman, A. Poznyak, "Observation & identification of mechanical systems via second order sliding modes", *Proceedings of the International Workshop on Variable Structure Systems*, Alghero, Italy, 2006, pp. 232-237.
- [19] A. Levant, "Quasi continuous high order sliding mode controllers", *IEEE Transactions on Automatic Control*, Vol. 50, No. 11, 2005, pp. 1812-1816.

Design of High-Speed Dual Port 8T SRAM Cell with Simultaneous and Parallel READ-WRITE Feature

Original Scientific Paper

Shourin Rahman Aura

Lecturer, Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
aura.eee@aust.edu

S. M. Ishraqul Huq

Assistant Professor, Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
Ishraqulhuq.eee@aust.edu

Satyendra N. Biswas

Professor, Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
sbiswas.eee@aust.edu

Abstract – An innovative 8 transistor (8T) static random access memory (SRAM) architecture with a simple and reliable read operation is presented in this study. LTspice software is used to implement the suggested topology in the 16nm predictive technology model (PTM). Investigations into and comparisons with conventional 6T, 8T, 9T, and 10T SRAM cells have been made regarding read and write operations' delay and power consumption as well as power delay product (PDP). The simulation outcomes show that the suggested design offers the fastest read operation and PDP optimization overall. Compared to the current 6T and 9T topologies, the noise margin is also enhanced. Finally, the comparison of the figure of merit (FoM) indicates the best efficiency of the proposed design.

Keywords: SRAM, CMOS, dual port, figure of merit

1. INTRODUCTION

Low power and high-speed integrated circuits (ICs) are continually in demand for portable applications such as mobile phones and laptops since high power dissipation reduces the battery life of electronic devices [1, 2]. For its superior performance in terms of the aforementioned metrics, static random-access memory (SRAM) is typically favored over dynamic RAM (DRAM), which requires frequent refreshing, to be employed in cache memories [3, 4]. The growth of the IC industry has led to the aggressive scaling of transistors to increase package density for low chip-area requirements [5]. Consequently, the size of SRAM circuits has decreased significantly [6]. As a result, power dissipation has become a growing concern while designing SRAMs in nanometer technology [7]. The two major operations, read and write, drive the performance of an SRAM cell and, therefore, the factors taken into consideration for design are the dynamic power consumption and latency during these two operations [8]. Dynamic

power is the power consumption when the SRAM cell reads or writes any data, and is the major contributor to power consumption in a circuit [9].

A traditional SRAM cell consists of cross-coupled complementary metal-oxide semiconductor (CMOS) inverters, each storing complementary outputs [10]. The storage nodes are accessed using pass-transistors for write, and sometimes read, operations. Existing 6T [11], 8T [12], 9T [13], and 10T [14] SRAM cells require pre-charging of the bit-lines (BL), or read bit-line (RBL) in some cases, during read operation and based on the stored values, one of the bit-lines is discharged. This pre-charging process requires additional circuitry and increases the read delay of the SRAM cell. Moreover, the existing circuits do not allow simultaneous read and write operations. Researchers have also proposed a load-less 4T [15] SRAM circuit to reduce transistor count and power penalty. However, the cell is highly unstable as the voltage level of one of the storage nodes degrades over time. There is no pull-up transistor to charge and hold the storage

nodes to the logic high supply rail. Budhaditya et. al. [16] presented a 5T SRAM cell which eliminates one access transistor to reduce power consumption and chip area. It facilitates the read operation as only one bit-line is pre-charged, thus, reducing read power dissipation. Nevertheless, the 5T circuit lacks a separate read circuit.

This study presents a modified 8T SRAM cell using the 5T write circuit and a CMOS inverter for a separate and isolated read operation. The objective of the proposed circuit is improved read operation with a simultaneous read-write feature.

2. LITERATURE REVIEW

The conventional 6T SRAM cell consists of two CMOS inverters cross-linked with two pass transistors connected to complimentary bit-lines, as shown in Fig. 1. The gates of access transistors M3 and M4 are connected to the write-line (WL), which enables write and read operations to allow data to be written to or read from the storage nodes Q and Qb [11]. During the read operation, the two bit-lines are pre-charged and when WL is enabled, one of the bit-lines will discharge through the access transistor and pull-down transistor depending on the stored logic levels. Although two bit-lines are not required, they are frequently used to increase noise margins by providing both the signal and its inverse [17].

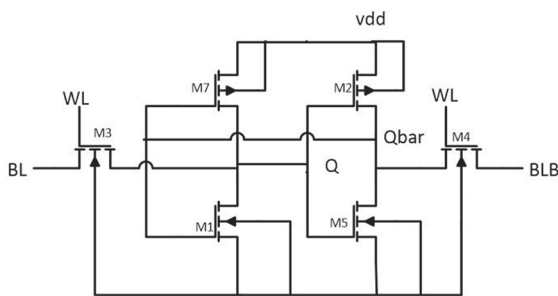


Fig. 1. Conventional 6T SRAM cell

The 6T SRAM circuit is the most area-efficient with a symmetrical layout compared to other SRAM designs [18]. However, the circuit requires pre-charging of the bit-lines and does not have separate write and read ports, thus, being unable to write and read simultaneously. For reliable operations, a certain cell-ratio (CR) and pull-up ratio (PR) needs to be maintained [19]. Process variation and supply scaling make the 6T architecture inefficient in the subthreshold region [20].

The 8T SRAM cell addresses several constraints of the 6T circuit including a separate read port, as shown in Fig. 2 [12]. The write operation is identical to that of the 6T circuit where the write word-line (WWL) is charged to turn on the access transistors and data is applied in the bit-lines. Data is stored in the nodes Q and Qb. During read operation, WWL is set to zero voltage and read bit-line (RBL) is pre-charged. When read word-line (RWL) is charged, RBL will discharge through M7 and M8 if '0' is stored in Q ('1' in Qb). When logic '1' is stored

in Q ('0' in Qb), the RBL will not discharge and remain at high logic. Although the write signal can be applied in this circuit during read function, the RBL will need to be recharged before the next read operation.

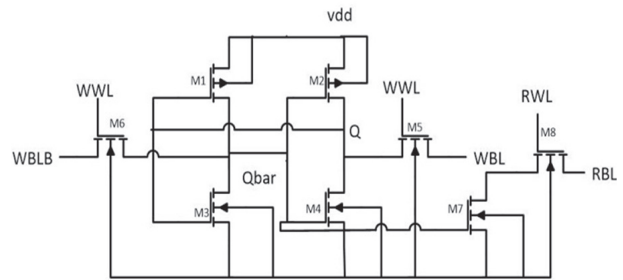


Fig. 2. 8T SRAM cell

The architecture of 9T SRAM cell consists of the 6T write cell with 3 additional transistors for the read circuit, as shown in Fig. 3 [13]. Write operation is identical to that of the 6T circuit. For the read operation, the bit-lines need to be pre-charged just like in the 6T SRAM cell. However, this time the discharging of the bit-lines occur through M7 or M8 (depending on the stored data of Q and Qb), and M9. The read-line (RL) signal enables the read circuit and if '1' is stored in Q, the BL is discharged through M7 and M9, while BLB remains at a high state. Since the storage nodes are isolated from the read path, the 9T SRAM cell provides an improved read static noise margin (RSNM) from the 6T SRAM cell.

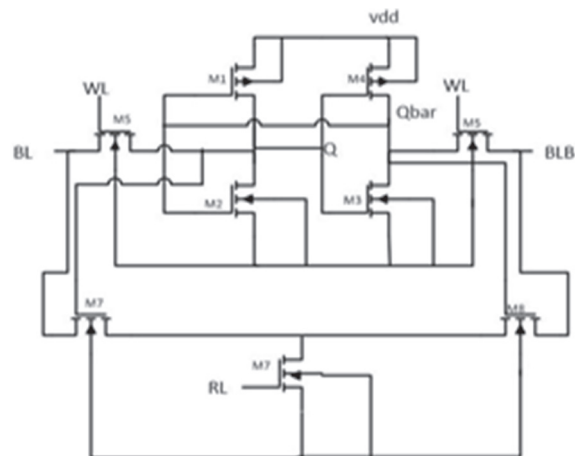


Fig. 3. 9T SRAM cell

Figure 4 shows a 10T SRAM cell presented in [14]. During the read operation, WL is charged while WWL remains at logic low, disabling transistors M3 and M4. Therefore, the storage nodes are isolated from the read path which improves the RSNM. The node VG (virtual ground) is set to 0V. The bit-lines are pre-charged and if '1' is stored in Q, then BL discharges through M8 and M9 while BLB remains at a high state. On the other hand, if '0' is stored in Q i.e. '1' stored in Qb, then BLB discharges through M6 and M10 while BL remains at the pre-charged value. The read operation is thus enabled by charging WL. During the write operation, both WL and WWL are charged, and VG is also

set to logic high voltage to disable the discharging path. The data from the bit-lines are, thus, stored at the storage nodes. However, since the write operation now includes two access transistors, the write delay is higher compared to other SRAM circuits.

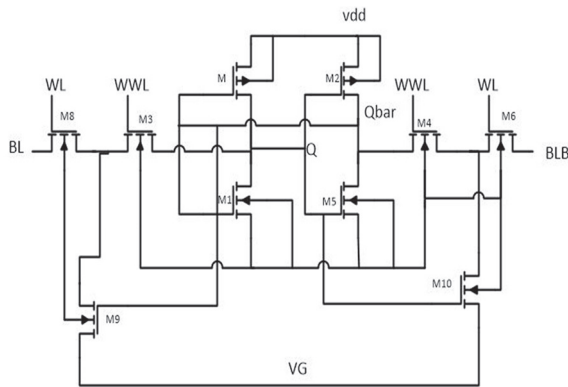


Fig. 4. 10T SRAM cell

3. PROPOSED DESIGN

The proposed 8T SRAM cell is shown in Fig. 5, which consists of a 5T write cell (M1-M5) and a 3T read cell (M6-M8). The write cell includes two cross-coupled CMOS inverters and an access transistor M3 which is enabled by the write-line (WL). Data is applied only through one bit-line and stored in the nodes Q and Qbar. The read cell consists of a single CMOS inverter and an access transistor M6 which is enabled by the read-line (RL). The CMOS inverter is connected between read-bit-line (RBL) and ground where RBL is charged during the read operation. Data is read from the node R/O. The access transistors M3 and M6 operate during write and read operations, respectively. A detailed operation of the SRAM cell is given below.

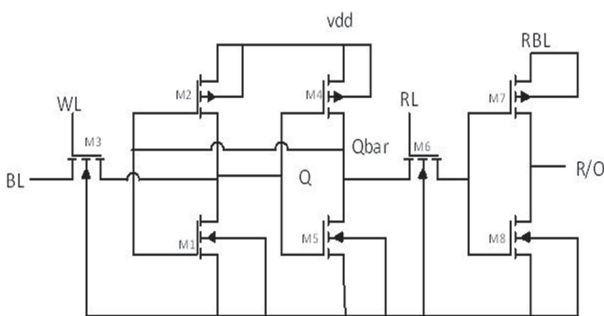


Fig. 5. Proposed 8T SRAM cell

Write '0': Write-line (WL) is charged and logic '0' is applied at BL. The data from BL is transferred through M3 to the storage node Q. The node Q is also connected to the gate of M4 which is turned on to pull-up Qbar to the supply voltage Vdd. As a result, logic '1' is stored in Qbar. This turns on the transistor M1 to connect node Q to the ground and, therefore, write '0'.

Write '1': Write-line (WL) is charged and logic '1' is applied at BL. The data from BL is transferred through M3 to the storage node Q. This high logic turns on M5 to

connect Qbar to the ground and store '0'. Logic '0' in node Qbar turns on M2 to connect Q to the supply voltage Vdd. After WL is disabled, the two storage nodes are latched and '1' is stored in node Q.

Read '0': During read operation, RBL is set to logic high voltage. When Q stores logic '0', '1' is stored in Qbar. After RL is charged, data '1' from Qbar is accessed which enables transistor M8 and the R/O is connected to the ground and '0' is read.

Read '1': When '1' is stored in Q, '0' is stored in Qbar. After RL is charged, logic '0' from Qbar is transferred through M6 to turn on M7 and connect R/O to RBL. Since RBL is set to logic high, '1' is read at the output.

Unlike the reported SRAM cells in literature, the proposed design circumvents the need of any pre-charging step during the read operation. The read circuit is completely isolated from the write circuit, thus, allowing simultaneous and parallel read and write operations. After RBL is charged, the read circuit is enabled and data can be read by charging RL. If RL is set to '0' and RBL remains at '1', previous data is read from the cell. During this scenario, new data can be written in parallel to the read function in the cell by charging WL while R/O keeps showing the previously stored data. Additionally, real-time data can be read simultaneously as it is being written if both RL and RBL remain at '1' during the write operation. However, when RBL is disabled, the output from R/O is invalid regardless of the state of RL. Furthermore, the isolation of the read path from storage nodes improves the read operation stability.

4. RESULTS

LTspice platform is used to simulate proposed and current SRAM cells in predictive technology model (PTM) 16 nm technology. Read and write latency, read and write power consumptions, PDP, and noise margin are evaluated to assess the performance of the topologies. The supply voltage was 0.9 V.

Figure 6 presents the transient operation of the proposed SRAM cell during write hold operations. With precise voltage levels, both '0' and '1' data were written successfully. Additionally, data is successfully overwritten without any distortions.

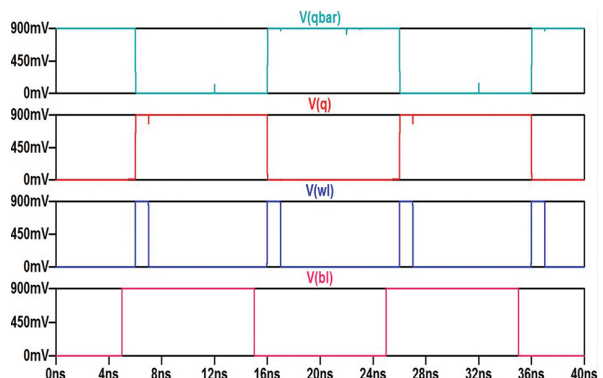


Fig. 6. Write operation of proposed 8T SRAM cell

It is considerably more difficult to read data in the existing SRAM architectures than to write it. The read data disappears once the read cycle is through. The circuit needs to be reset to read the data again. However, successive read operations can be easily performed by simply charging RL repeatedly. The voltage at R/O stays constant throughout subsequent read operations, which is evident from Fig. 7. Repetitive read operations do not add any initialization or propagation delay. RBL can be maintained at a logic low when a read operation is not required, which lowers the circuit's power requirement.

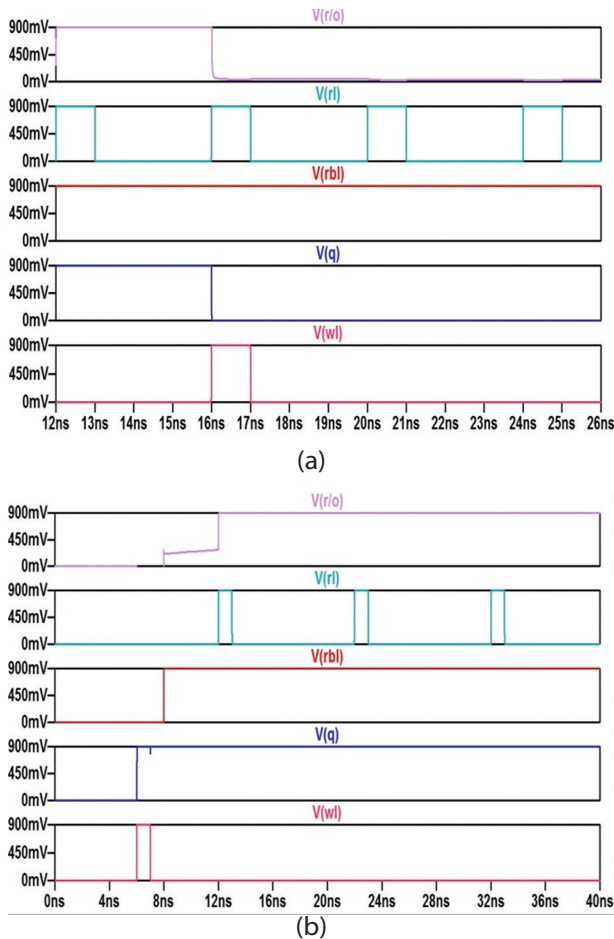


Fig. 7. (a) Successive read '0', and (b) successive read '1' operations of proposed 8T SRAM

Simultaneous read and write operations in the proposed SRAM cell are shown in Fig. 8. Once RL and RBL are charged to logic high, the read circuit is enabled and data is read at R/O. When WL is enabled to write data, R/O changes with the storage node Q, and data is read simultaneously as it is written. The write action is unaffected by the read operation.

Fig. 9 shows the comparison of write '0' and write '1' delay for all the studied SRAM cells at a supply voltage of 0.9 V. The proposed 8T SRAM cell has lower write '0' latency than all the other SRAM cells and lowers write '1' delay compared to 8T and 10T SRAM cells. Fig. 10 shows the comparison of the read '0' and read '1' delay for all the studied SRAM cells. The read operation for the existing circuits with pre-charging is performed by

adding a 1 fF capacitor. Based on the results, the proposed design significantly improves the read delay.

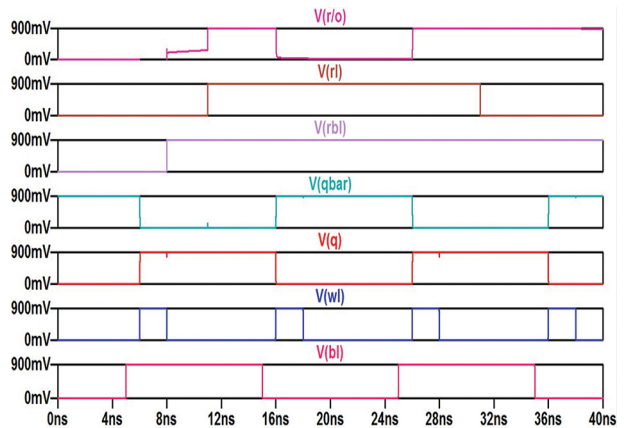


Fig. 8. Simultaneous Read-write operations of proposed 8T SRAM

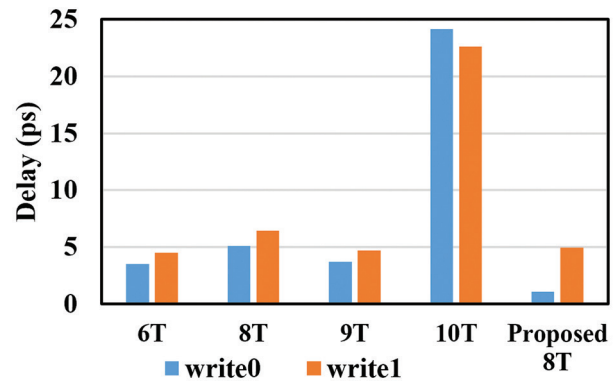


Fig. 9. Calculated write delay comparison

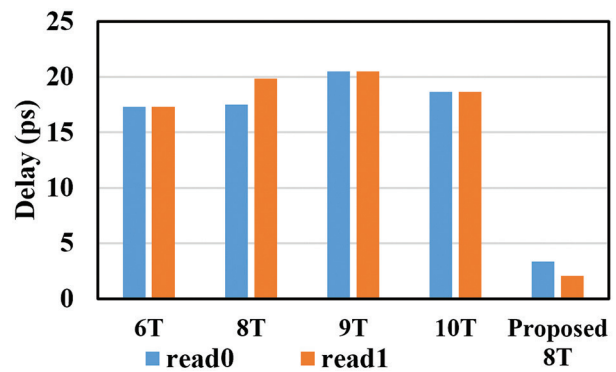


Fig. 10. Calculated read delay comparison

Table 1. Comparison of write and read power of different SRAM cells

Cell	Power (uW)			
	Write 0	Write 1	Read 0	Read 1
6T	12.56	12.56	0.93	0.92
8T	11.75	11.05	0.22	0.34
9T	13.06	12.70	1.40	1.40
10T	13.53	13.53	0.22	0.22
Proposed 8T	18.09	5.19	0.99	0.52

The write and read power consumptions of the investigated cells are compared in Table 1. Results show that the proposed 8T SRAM cell has a lower power penalty for write '1' operation compared to previously reported topologies, but a higher write '0' and comparable read power penalties. The increased power consumption during write '0' is due to the absence of the complementary bit-line and voltage drop across the single access transistor when passing logic '0'.

Fig. 11 presents the comparison of PDP of different SRAM topologies. Our proposed 8T SRAM cell has the lowest write '1' PDP compared to other architectures. For read '1' operation the proposed cell has the lowest PDP, while read '0' PDP is significantly lower compared to that of 6T and 9T cells.

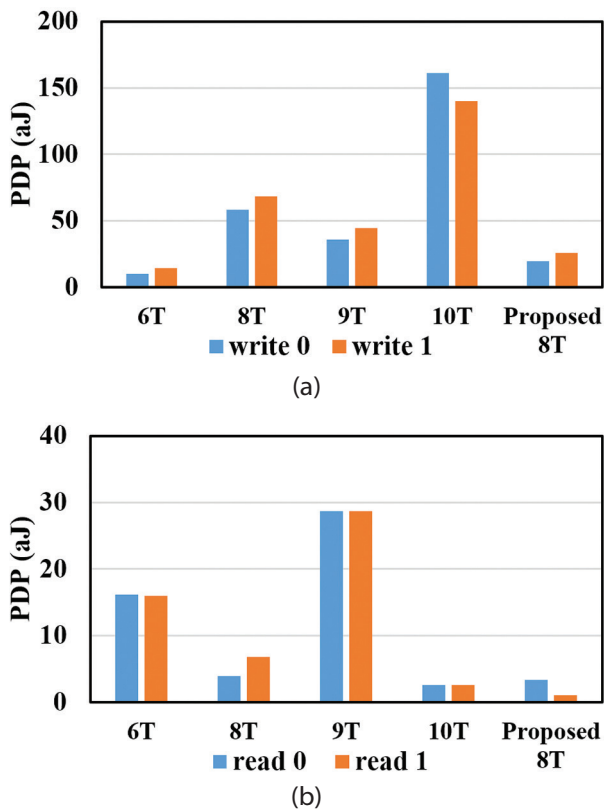


Fig. 11. PDP of (a) write, and (b) read operations of different SRAM topologies.

The noise margin of the proposed cell is compared with existing 6T, 8T and 9T topologies in Fig. 12. Results show that the proposed read circuit provides an improved read noise margin compared to 6T and 9T cells. Moreover, the proposed cell has the largest write '0' noise margin.

Since different SRAM topologies may perform superior to others in terms of different parameters, it is necessary to compare them in terms of a single numerical quantity that considers all other characteristics to evaluate the efficiency. This parameter is called the figure of merit (FoM) and is defined by Equation (1), where a larger value indicates better performance. Table 2 compares the FoM of different topologies and the results

show that the proposed SRAM cell is the most efficient with the highest value.

$$FoM = \frac{\text{sum of Noise Margins}}{\text{sum of PDPs}}$$

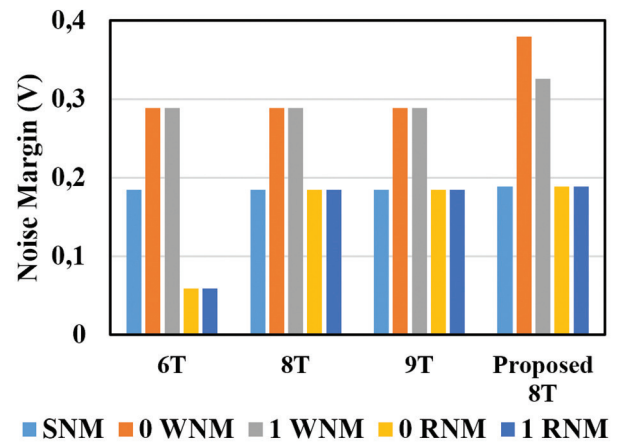


Fig. 12. Comparison of Noise Margin of Proposed cell with the other SRAM topologies.

Table 2. FoM comparison of different SRAM cells

Cell	FoM
6T	15.50
8T	8.24
9T	8.23
Proposed 8T	25.60

The proposed cell is also simulated using the 10nm Berkeley Short-channel IGFET Model-Common Multi-Gate (BSIM-CMG) FinFET model to investigate the robustness of the circuit at a different technology node. The performance is compared with the conventional 6T SRAM cell. Tables 3, 4, and 5 present the comparisons for delay, power, and PDP, respectively. The proposed cell has a lower read and write delay. Although power consumption is higher, the PDP is significantly lower for write and read '0' and read '1'.

Table 3. Comparison of write and read delay of different SRAM cells at 10nm node

Cell	Delay (ps)			
	Write 0	Write 1	Read 0	Read 1
6T	2.99	3.99	22.8	22.8
Proposed 8T	0.58	1.26	3.64	1.84

Table 4. Comparison of write and read power of different SRAM cells at 10nm node

Cell	Power (uW)			
	Write 0	Write 1	Read 0	Read 1
6T	9.48	9.48	0.10	0.10
Proposed 8T	13.8	0.60	1.6	0.94

Table 5. Comparison of write and read PDP of different SRAM cells at 10nm node

Cell	PDP (aJ)			
	Write 0	Write 1	Read 0	Read 1
6T	28.35	37.83	2.30	2.30
Proposed 8T	8.00	0.75	5.82	1.72

5. CONCLUSIONS

This study compares a modified SRAM cell with 8 transistors to previously reported standard cells, such as 6T, 8T, 9T, and 10T. Making reading operations more effective and simple is the aim of this study. The recommended SRAM cell separates the read and write operations and does away with the necessity for bit line pre-charging. As a result, it is simple and reliable to carry out simultaneous read and write operations. The read 0 delay is reduced by 80.70%, 80.91%, 83.70%, and 82.07% when compared to 6T, 8T, 9T, and 10T cells, respectively. Meanwhile, read 1 latency is reduced by 88.16 percent, 89.68 percent, 90 percent, and 89 percent, respectively, in comparison to 6T, 8T, 9T, and 10T cells, in addition to enhanced noise margin and low PDP. However, the difference in delay for read '0' and read '1' indicates unsymmetrical operation which may be improved along with power consumption in future work.

6. REFERENCES

- [1] Y. He, J. Zhang, X. Wu, X. Si, S. Zhen, B. Zhang, "A half-select disturb-free 11T SRAM cell with built-in write/read-assist scheme for ultralow-voltage operations", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 27, No. 10, 2019, pp. 2344–2353.
- [2] S. Gupta, K. Gupta, N. Pandey, "Pentavariate Vmin analysis of a subthreshold 10T SRAM bit cell with variation tolerant write and divided bit-line read", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 65, No. 10, 2018, pp. 3326–3337.
- [3] A. Agal, Pardeep, B. Krishan, "6T SRAM Cell: Design And Analysis", *International Journal of Engineering Research and Applications*, Vol. 4, No. 3, 2014, pp. 574–577.
- [4] M. Devi, C. Madhu, N. Garg, "Design and analysis of CMOS based 6T SRAM cell at different technology nodes", *Materialstoday: proceedings*, Vol. 28, No. 3, 2020, pp 1695-1700.
- [5] V. Sharma, F. Catthoor, W. Dehaene, "SRAM bit cell optimization," *SRAM Design for Wireless Sensor Networks*, Springer New York, 2013, pp. 9-30.
- [6] D. Sharma, S. Birla, "Design and Analysis of 10T SRAM Cell with Stability Characterizations", *Proceedings of the International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies*, Bhilai, India, 19-20 February 2021.
- [7] W. Lim, H. C. Chin, L. S. Cheng, M. L. P Tan, "Performance evaluation of 14nm FinFET-based 6T SRAM cell functionality for DC and transient circuit analysis", *Journal of Nanomaterials*, Vol. 2014, July 2014.
- [8] A. Azizi-Mazreah, M. T. M. Shalmani, H. Barati, A. Barati, "Delay and energy consumption analysis of conventional SRAM", *International Journal of Electrical and Computer Engineering*, Vol. 2, No. 1, 2008, pp. 35-39.
- [9] Y. Kumar, S. K. Kingra, "Stability analysis of 6T SRAM cell at 90nm technology", *International Journal of Computer Applications*, 2016, pp. 32-36.
- [10] M. Yamaoka, Y. Shinozaki, N. Maeda, Y. Shimazaki, K. Kato, S. Shimada, K. Yanagisawa, K. Osada, "A 300-MHz, 25 μ A /Mbitleakage on-chip SRAM module featuring process-variation immunity and low-leakage-active mode for mobile-phone application processor", *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 1, 2005, pp. 186-194.
- [11] N. Rahman, B. Prasad Singh, "Static-Noise-Margin Analysis of Conventional 6T SRAM Cell at 45nm Technology", *International Journal of Computer Applications*, vol. 66, No. 20, 2013, pp. 19-23.
- [12] D. Mittal, V. K. Tomar, "Performance Evaluation of 6T, 7T, 8T, and 9T SRAM cell Topologies at 90 nm Technology Node", *Proceedings of the 11th International Conference on Computing, Communication and Networking Technologies*, Kharagpur, India, 1-3 July 2020, pp. 1-4.
- [13] C. Yu, M. Shiao, "Single-Port Five-Transistor SRAM Cell with Reduced Leakage Current in Standby", *International Journal of VLSI Design & Communication Systems*, Vol. 7, No. 4, 2016, pp. 1-11.
- [14] G. Prasad, "Novel low power 10T SRAM cell on 90nm CMOS", *Proceedings of the 2nd International Conference on Advances in Electrical, Electronics,*

Information, Communication and Bio-Informatics, Chennai, India, 27-28 February 2016, pp. 109-114.

- [15] J. Yang, L. Chen, "A New Loadless 4-Transistor SRAM Cell with a 0.18 μm CMOS Technology", Proceedings of the Canadian Conference on Electrical and Computer Engineering, Vancouver, BC, Canada, 22-26 April 2007, pp. 538-541.
- [16] B. Majumdar, S. Basu, "Low Power Single Bitline 6T SRAM Cell With High Read Stability", Proceedings of the international Conference on Recent Trends in Information Systems, Kolkata, India, 21-23 December 2011, pp. 169-174.
- [17] A. S. V. S. V. P. D. Kumar, B. S. Suman, C. A. Sarkar, D. V. Kushwaha, "Stability and Performance Analysis of Low Power 6T SRAM Cell and Memristor Based SRAM Cell using 45NM CMOS Technology", Proceedings of the International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering, Bhubaneswar, India, 27-28 July 2018, pp. 2218-2222.
- [18] K. Osada, Y. Saitoh, E. Ibe, K. Ishibashi, "16.7-fA/cell tunnelleakage-suppressed 16-Mb SRAM for handling cosmic-ray-induced multierrors", Proceedings of the IEEE International Solid-State Circuits Conference, San Francisco, CA, USA, 13 February 2003, pp. 302-494.
- [19] H. Kumar, V. K. Tomar, "A Review on Performance Evaluation of Different Low Power SRAM Cells in Nano-Scale Era", Wireless Personal Communication, Vol. 117, 2020, pp. 1959-1984.
- [20] M. U. Mohammed, A. Nizam, Liaquat Ali, M. H. Chowdhury, "FinFET based SRAMs in Sub-10nm domain", Microelectronics Journal, Vol. 114, 2021, pp. 1-14.

Comparative Performance of DVR and STATCOM for Voltage Regulation in Radial Microgrid with High Penetration of RES

Case Study

Ritika Gour

Delhi Technological University,
Electrical Engineering Department, Delhi, India
riti113@gmail.com

Vishal Verma

Delhi Technological University,
Electrical Engineering Department, Delhi, India

Abstract – In recent years, the penetration of renewable energy sources (RES) in microgrids and distribution system feeders has increased manifold. Moreover, the advancement of power electronics-based devices in the distribution system has significantly increased the number of sensitive loads. Variations in the voltage under intermittent RES create functional problems with sensitive loads, necessitating voltage regulators (VR) installation. In this paper, two custom power devices: dynamic voltage restorer (DVR) and static synchronous compensator (STATCOM), used for voltage regulation in a microgrid, are investigated under different operating conditions. The efficacy of DVR and STATCOM for voltage regulation in an 11-node radial microgrid with high penetration of RES is simulated under a MATLAB Simulink environment. Furthermore, the simulated microgrid voltage profile results are analyzed to evaluate the efficacy of both voltage regulators.

Keywords: RES, intermittency, CPL, microgrid, voltage regulation, DVR, STATCOM

1. INTRODUCTION

A microgrid is a local energy grid with a group of connected energy sources and loads that usually operate in synchronization with the conventional grid but can also operate independently in the event of any anomaly. Various renewable energy sources (RES) can be connected to the microgrid such as solar panels, wind turbines, etc. With the advancement of technology, the diminishing supply of conventional power sources and many environmental and socio-economic factors have raised RES penetration in the microgrid [1-4]. Increased renewable energy in the microgrid provides numerous benefits, like increased local power availability, low-cost, clean energy, increased reliability and resilience, reduced grid congestion and peak loads, etc. However, the intermittent nature of RES creates power fluctuations and large variations in the voltage profile. As a result of the unpredictable nature of these RES, the microgrid operator is unable to schedule the load, posing a risk to the entire system. Furthermore, the widespread use of power electronics-based devices in residential and commercial loads has significantly increased the number of sensitive loads. Power and voltage variations due to high penetration of RES may result in the maloperation of these sensi-

tive loads, damage to equipment, and cascading faults. These challenges deteriorate the power quality of the microgrids which causes instability and monetary losses for both the microgrid and the consumer. [5].

The variability of demands and intermittency of renewable sources necessitates the methods to deal with variation in the voltage of the microgrid. Mitigation of the negative effects of intermittent RES is achieved by adopting different methods for regulating the voltage of the microgrid. The static synchronous compensator (STATCOM) and the dynamic voltage restorer (DVR) are two custom power devices that are generally installed in the distribution system to regulate voltage profile [8-10]. DVR regulates the voltage at the point of common coupling (PCC) by supplying/absorbing the voltage in series with the PCC voltage and is connected to the feeder through an injection transformer [5] [9-12]. On the other hand, STATCOM is connected in shunt with the feeder, thereby allowing the supply or absorption of current from the feeder, with an effect on the voltage as well [10] [13].

The efficacy of DVR and STATCOM for voltage regulation in microgrids with significant RES penetration is in-

investigated in this paper by simulating an 11-node grid-connected radial microgrid in the MATLAB/Simulink environment. Findings of the simulation are then used to draw a microgrid voltage profile from the grid mains to the last node in the presence of the STATCOM and DVR, one by one, and their efficacy is assessed.

Section 2 describes the block diagram of the considered system, and Section 3 discusses the block diagram and phasor diagram of both voltage regulators DVR and STATCOM. Further Section 4 includes MATLAB Simulink and performance evaluations of both STATCOM and DVR.

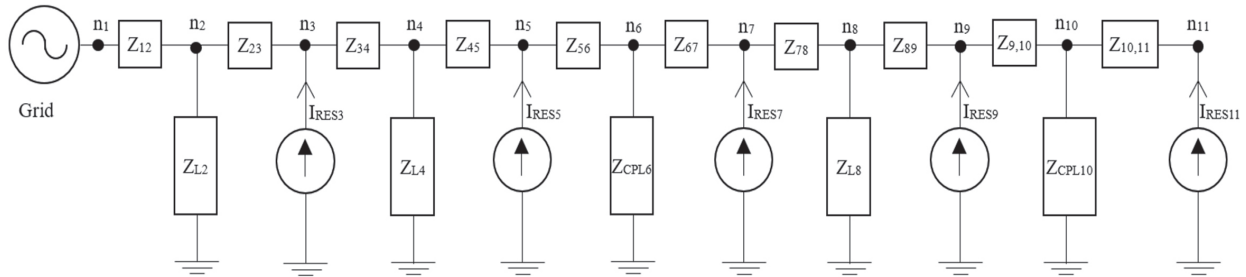


Fig. 1. Single line diagram of system under consideration

2. SYSTEM CONFIGURATION

The system under consideration is shown in Fig. 1. Considered a system an 11-node grid connected radial microgrid. The voltage source/grid is assumed to be a reference node (n_1). RES sources (I_{RES_i}) and loads Z_{L_i} are connected alternatively throughout the radial microgrid (where 'i' is the node number with respect to the reference node). All the RESs are connected on odd nodes (n_3, n_5, n_7, n_9 and n_{11}) and all the loads are connected on even nodes (n_2, n_4, n_6, n_8 and n_{10}) with respect to the reference node (n_0). In the considered system RESs are connected to the microgrid as a current source feeding the microgrid through a current-controlled voltage source converter (VSC). Loads connected to the microgrid are of two types: constant impedance load (CIL) and constant power load (CPL). CILs are connected at node 2, node 4, and node 8. Generalized expression for the impedance Z_{L_i} of the CIL is given as:

$$Z_{L_i} = \frac{(V_{n_i}(\text{rated}))^2}{P_{L_i} + jQ_{L_i}} \quad (1)$$

Where $V_{n_i}(\text{rated})$ is rated voltage at the node and $P_{L_i} + jQ_{L_i}$ is rated power of the load (power consumed by the load at rated voltage).

CPLs are connected at node 6 and node 10. The impedance of CPL is dynamic in nature, it shows a negative impedance characteristic, and power drawn (P_{CPL_i}) remains constant irrespective of the instantaneous voltage (V_{n_i}) at its terminal. CPL changes load current according to the voltage level so that it can draw a constant power from the microgrid. Generalized expression for the dynamic impedance of the CPL is given as [14-15]:

$$Z_{CPL_i} = \frac{(V_{n_i})^2}{P_{CPL_i}} \quad (2)$$

For ease of study, certain assumptions are made which are as follows:

- RES and loads are placed alternatively on the microgrid with uniform distancing.
- Microgrid is part of distribution feeder with R/X ratio ≈ 8 such that feeder impedance is $0.642 + j0.0833 \Omega/\text{km}$ [16].
- Length of each section is 0.5 km.

3. OPERATING PRINCIPLE OF DVR AND STATCOM

3.1 OPERATING PRINCIPLE OF DVR

DVR is a custom power device connected in series with the feeder as seen in Fig. 2(a). DVR generally comprises of injection transformer, ripple filter, VSC and DC link. The DC voltage on the DC link is converted to AC by VSC as required for voltage regulation. A further ripple filter is used to filter the AC voltage generated from VSC before the injection transformer injects it into the line.

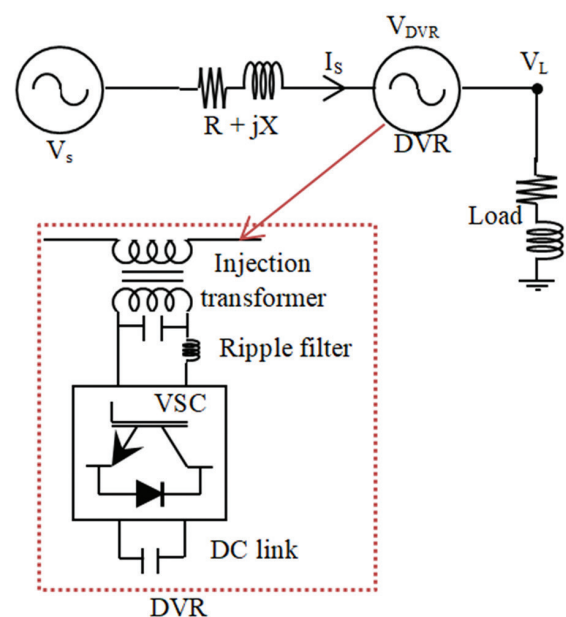


Fig. 2.(a) Block diagram for voltage regulation by DVR

The injection transformer is a low leakage impedance transformer that connects the DVR to the feeder in series. On the DC side of the VSC, a capacitor is typically present to maintain the DC link voltage. To increase the range of regulation of DVR, batteries can also be placed on the DC side. The phasor diagram for the DVR voltage regulation is shown in Fig. 2(b). From the phasor diagram, it can be observed that the net voltage at the terminal of DVR is equal to the phasor sum of the previous terminal voltage and DVR injected/absorbed voltage.

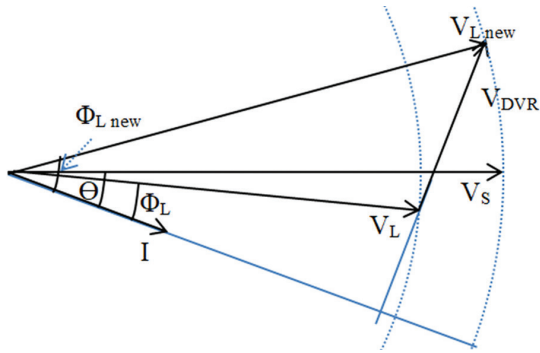


Fig. 2.(b) Phasor diagram of DVR for voltage regulation

Synchronous reference frame theory is used to control the gating pulses of VSC in DVR, which are subsequently used for controlling the injection of requisite voltage from the DVR [9]. Fig. 2(c) shows the block diagram of the control scheme.

The difference between the measured voltage, from the terminals of DVR, and the rated voltage is passed through the proportional-integral (PI) controller for generating the DVR reference (injection/absorption) voltage. Usually, the voltage regulation is done with the reactive component of voltage, which means that the DVR reference voltage is generally the quadrature-axis (q-axis) component (V_q). Nevertheless, if the voltage required to regulate the terminal voltage exceeds the limit of the DVR's reactive capacity, then V_q is reduced with a simultaneous increase of V_d to regulate the terminal voltage of the DVR. Subsequently, reverse park transformation is performed on the reference V_d and V_q through dq to abc transformation. The gating pulses of VSC are generated by passing the resultant abc reference signal through the pulse width modulator (PWM) block. These gating pulses are further used to control the ON/OFF different switches of VSC, for requisite voltage injection, to regulate the voltage at the DVR terminal.

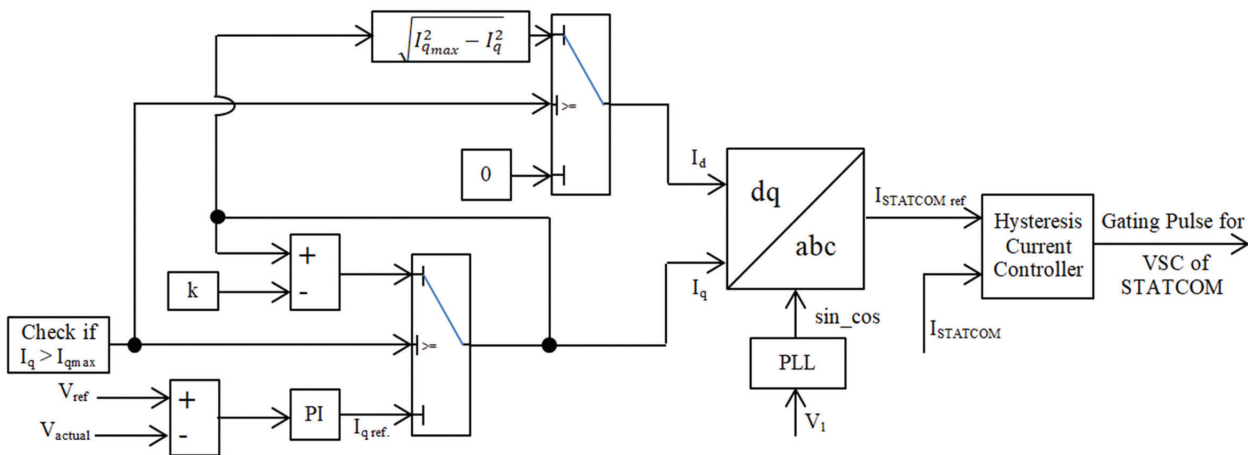


Fig. 2. (c) Block diagram of control of STATCOM

3.2 OPERATING PRINCIPLE OF STATCOM

Another custom power device investigated in the paper is STATCOM, which is in shunt with the feeder, as illustrated in Fig. 3(a). The VSC converts DC to AC, and after passing through the ripple filter, the output current from the VSC is sent to the microgrid. In the design shown in Fig. 3(a), it is connected in parallel to the load connected at nodes 6 and 10. Fig. 3(b) shows a phasor diagram for its voltage regulation.

The synchronous reference frame theory is also used to control the VSC of STATCOM. Fig. 3(c) depicts the control approach for generating the gating pulses for the VSC of STATCOM. The controller senses the volt-

age at its terminal, and then the difference between the sensed voltage and the reference voltage is passed through the PI controller, generating the q-axis component of the current (I_q). As shown in the block diagram, when the STATCOM approaches its reactive power limit, the q-axis component is reduced, and the d-axis component (I_d) is increased, as done in the case of DVR. To generate the reference STATCOM current ($I_{STATCOMref}$), the reference values of I_d and I_q are transformed into reference abc signal through reverse PARKs transformation. Furthermore, the gating pulses for VSC switches are generated by passing STATCOM reference ($I_{STATCOMref}$) current and actual ($I_{STATCOM}$) current through a hysteresis current controller.

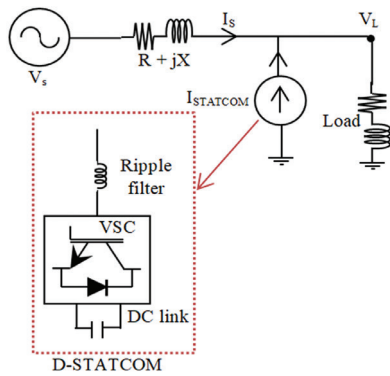


Fig. 3.(a) Block diagram for voltage regulation by STATCOM

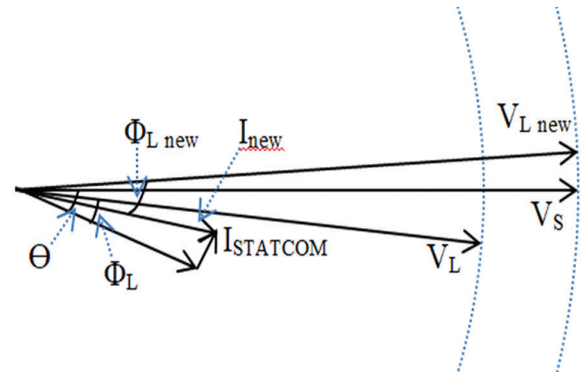


Fig. 3.(b) Phasor diagram of STATCOM for voltage regulation

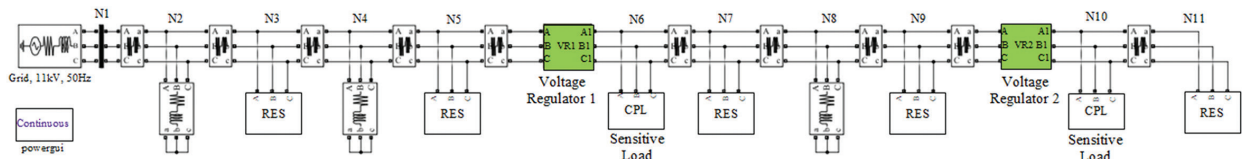


Fig. 4.(a) MATLAB/Simulink diagram for considered radial microgrid

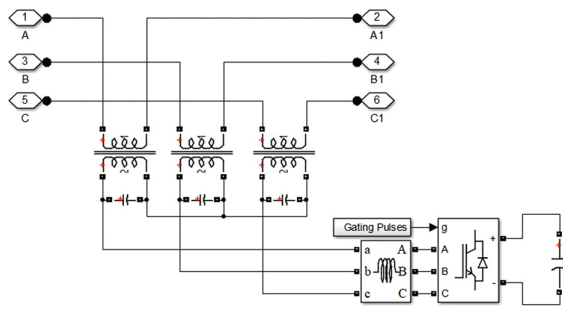


Fig. 4.(b) MATLAB/Simulink diagram for DVR

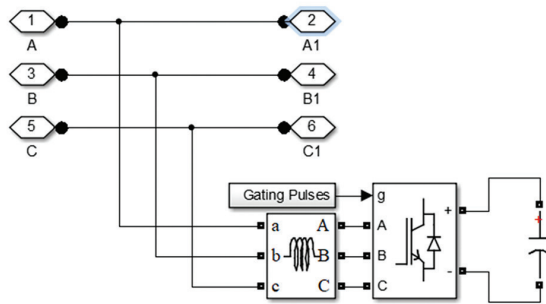


Fig. 4.(c) MATLAB/Simulink diagram for STATCOM

4. PERFORMANCE EVALUATION OF DVR AND STATCOM

Simulations in the MATLAB/Simulink environment are done to investigate the efficacy of DVR and STATCOM in the considered 11 nodes radial microgrid. MATLAB simulation diagram of the microgrid with RES and voltage regulator is shown in Fig. 4(a). The voltage regulator can be either DVR or STATCOM and are connected at node 6 and node 10 and the expanded figures are shown in Fig. 4(b) and Fig. 4(c) for DVR and

STATCOM respectively. Parameters considered for the simulation of the above-mentioned configuration are listed in Table 1.

Table 1. Parameters used for simulation of the considered system

Parameters	Values
Source/Grid Voltage	11 kV, 50 Hz
Feeder impedance	0.642 + j 0.083 Ω/km, Length of each section = 0.5 km
RES rating	0.75 MW for each unit of RES
CPL rating	1 MW for each CPL load
Constant Impedance load (CIL)	0.5 MVA each load, R = 193 Ω, L = 462 mH
DVR	0.5 MVA for each unit.
STATCOM	0.5 MVA for each unit.

Both the STATCOM and the DVR are assigned the same rating to evaluate their regulation proficiency. The effect of voltage control is reported for different levels of RES penetration: (i) RES penetration is high with perturbing loads ($I_{RES} = 60$ A). (ii) RES penetration is low with perturbing loads ($I_{RES} = 25$ A).

(i) High penetration of RES

In this case, high penetration of RES is considered, as each RES is generating maximum power as per its rating. In this scenario, the RES current $I_{RES} = 60$ A (peak), resulting in each unit contributing 0.75 MW of power to the microgrid. The voltage profile is used to assess the responsiveness of DVR and STATCOM for voltage regulation in this scenario for several loading conditions. All the graphs, for each condition, show the voltage profile of the microgrid with considered loading conditions for:

- i. without any RES and voltage regulator connected to it.
- ii. with RES and without any voltage regulator connected to it.
- iii. with RES and DVR as the voltage regulator.
- iv. with RES and STATCOM as voltage regulators so that the regulation is accomplished by real current injection.
- v. with RES and STATCOM as voltage regulators so that the regulation is accomplished by reactive current injection.

a. High penetration of RES and rated loading condition:

The voltage profile of the microgrid with high penetration of RES at rated loading conditions, with each CIL drawing 0.5 MVA and each CPL drawing 1 MW power, is shown in Fig. 5(a). The voltage profile reveals that both the STATCOM and the DVR can regulate the voltage at their point of common coupling (PCC), i.e., at node 6 and node 10. DVR can regulate voltage with reactive power injection (V_{qDVR}), whereas STATCOM ($I_{dSTATCOM}$) can do the same with real power injection. Furthermore, Fig. 5(a) depicts the STATCOM's ($I_{qSTATCOM}$) voltage profile while it is regulating through reactive power, yet the voltage does not reach the rated value even at the point of common coupling, but the profile is improved slightly.

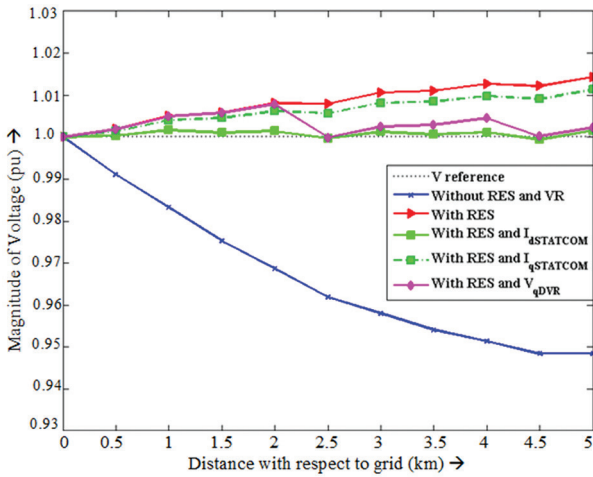


Fig. 5.(a) Voltage profile of microgrid with high RES penetration and rated loading condition

b. High penetration of RES and light loading condition:

The voltage profile of the microgrid with high penetration of RES and light loading conditions, i.e., 20% of the rated value, can be seen in Fig. 5(b). In this instance, the DVR improves the overall voltage profile of the microgrid by regulating the voltage at its two connecting points by absorbing a voltage that is a combination of both real and reactive power. However, since the range

required in this instance is relatively large, the STATCOM appears to be inadequate in regulating the voltage using either real or reactive current transactions.

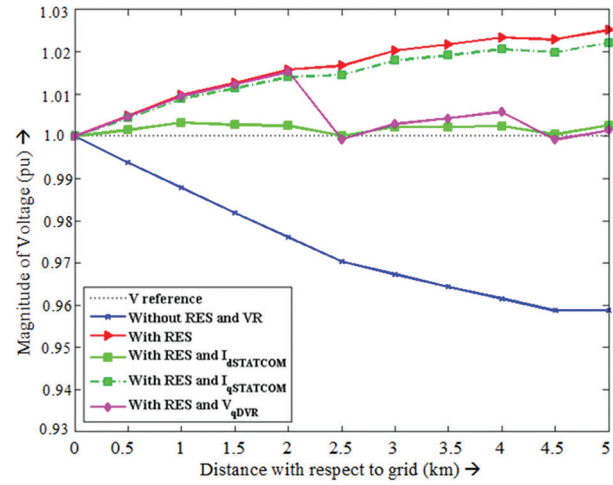


Fig. 5.(b) Voltage profile of microgrid with high RES penetration and light loading condition

c. High penetration of RES and heavy loading conditions:

Figure 5(c) depicts the voltage profile of a microgrid with high-RES penetration and heavy loading conditions of about 150% of the rated value. The voltage profile of the microgrid is enhanced with higher injection from RES because the loads are quite high.

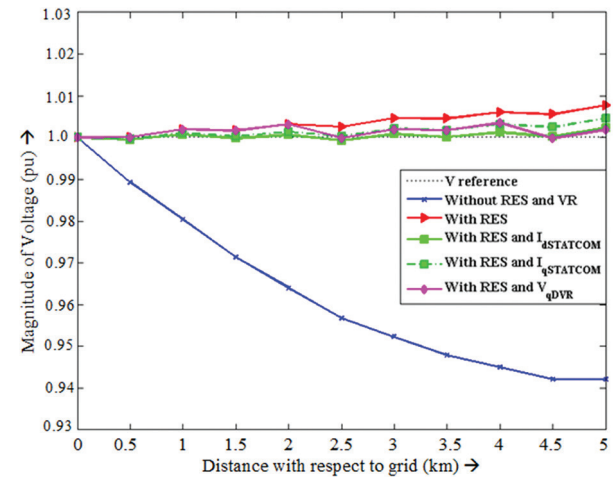


Fig. 5.(c) Voltage profile of microgrid with high RES penetration and heavy loading condition

As a result, both DVR and STATCOM improve the voltage profile even further.

(ii) Low penetration of RES and normal loading condition:

Injection from all the RES is reduced to 40% of the maximum value (given in the preceding scenario), with $I_{RES} = 20$ A (rms), and each unit now supplies 0.30 MW of power to the microgrid.

a. Low penetration of RES and normal loading condition:

Figure 5(d) depicts the voltage profile of the microgrid under low RES penetration and rated load conditions. In this scenario, the DVR utilizes reactive voltage injection to regulate the voltage at its terminals, but the STATCOM employs real current injection.

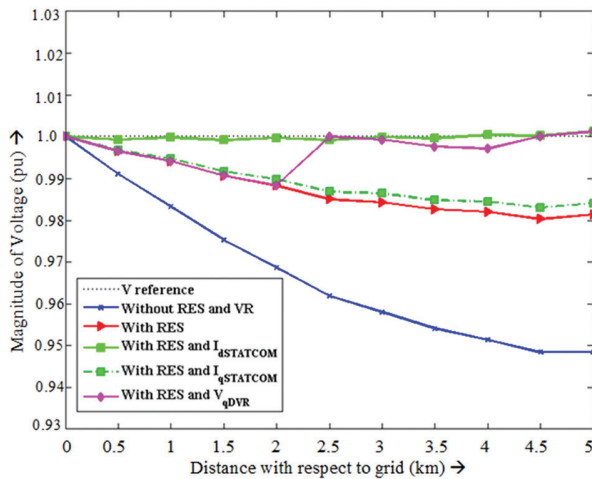


Fig. 5.(d) Voltage profile of microgrid with low RES penetration and rated loading condition

b. Low penetration of RES and light loading condition:

Figure 5(e) depicts the voltage profile of the microgrid at low RES penetration and light loading conditions, i.e. 20% of the rated value. As soon as the RES starts supplying power to the microgrid, the voltage profile improves. The voltage profile is further improved by both DVR and STATCOM.

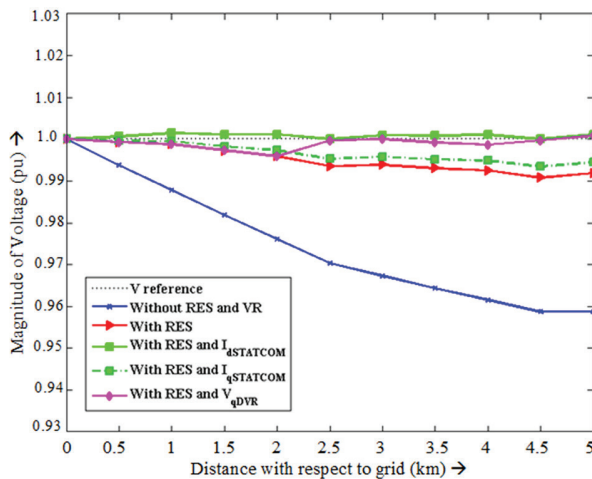


Fig. 5.(e) Voltage profile of microgrid with low RES penetration and light loading condition

c. Low penetration of RES and heavy loading condition:

Figure 5(f) depicts the voltage profile of the microgrid at low RES penetration and heavy loading conditions,

i.e. 150% of the rated value. Because the voltage variation is large in this scenario, the DVR requires both real and reactive power to perform, but it regulates the voltage, whereas the STATCOM is unable to do so even with both the real and reactive current injection.

5. CONCLUSION

The efficacy of DVR and STATCOM has been investigated and presented in this paper for improving the voltage profile of the 11-node radial microgrid with intermittent RES source in the MATLAB simulation environment. The PI controller and synchronous reference frame theory are used to control STATCOM and DVR. Voltage profiles obtained from the simulations demonstrate the comparison of the efficacy of DVR and STATCOM for voltage regulation in the microgrid. Under significant voltage variations, STATCOM can regulate the voltage only with real power support (through battery), while DVR allows voltage regulation with reactive power unless the voltage variations are extremely high. Utilization of STATCOM as a voltage regulator significantly increases the cost of the system as battery cost is also added to the system. In conclusion, the STATCOM is appropriate for voltage regulation if the range of compensation required is small, but as the range of compensation required widens, the DVR appears to be a superior solution.

6. REFERENCES

- [1] D. Ma, M. Liu, H. Zhang, R. Wang, X. Xie, "Accurate Power Sharing and Voltage Regulation for AC Microgrids: An Event-Triggered Coordinated Control Approach", IEEE Transactions on Cybernetics, 2021, pp. 1-11.
- [2] J. von Appen, M. Braun, T. Stetz, K. Diwold, D. Geibel, "Time in the Sun: The Challenge of High PV Penetration in the German Electric Grid", IEEE Power and Energy Magazine, Vol. 11, No. 2, 2013, pp. 55-64.
- [3] N. S. Jayalakshmi, P. B. Nempu, "Performance Enhancement of a Hybrid AC-DC Microgrid Operating with Alternative Energy Sources Using Supercapacitor", International Journal of Electrical and Computer Engineering Systems, Vol. 12 No. 2, 2021.
- [4] T. V. Krishna, M. K. Maharana, C. K. Panigrahi, "Integrated Design and Control of Renewable Energy Sources for Energy Management", Engineering, Technology & Applied Science Research, Vol. 10, No. 3, 2020, pp. 5857-5863.
- [5] V. Verma, R. Gour, "OLTC-DVR hybrid for voltage regulation and averting reverse power flow in the

micro-grid with intermittent renewable energy sources", Proceedings of the IEEE Industrial Electronics and Applications Conference, Kota Kinabalu, Malaysia, 2016, pp. 81-87.

- [6] Y. He, M. Wang, Z. Xu, "Enhanced Voltage Regulation of AC Microgrids with Electric Springs", Proceedings of the IEEE Applied Power Electronics Conference and Exposition, Anaheim, CA, USA, 17-21 March 2019, pp. 534-539.
- [7] A. Pimenta, P. B. C. Costa, G. M. Paraíso, S. F. Pinto, J. F. Silva, "Active Voltage Regulation Transformer for AC Microgrids", Proceedings of the IEEE 9th International Power Electronics and Motion Control Conference, Nanjing, China, 2021, pp. 2012-2017.
- [8] A. Ghosh, G. Ledwich, "Power Quality Enhancement Using Custom Power Devices", The Springer International Series in Engineering and Computer Science book series, 2002, pp 113-136.
- [9] R. Gour, V. Verma, "Voltage Regulation in a Radial Microgrid with High RES Penetration: Approach-Optimum DVR Control", Engineering, Technology & Applied Science Research, Vol. 12, No. 4, 2022, pp. 8796-8802.
- [10] H. M. A. Rashid, S. A. Jumaat, S. H. N. Yusof, S. A. Zulkifli, "Modeling the Grid Connected Solar PV (GCPV) System with D-STATCOM to Improve Stability System", Proceedings of the IEEE International Conference in Power Engineering Application, Shah Alam, Malaysia, 7-8 March 2022, pp. 1-6.
- [11] A. H. Soomro, A. S. Larik, M. A. Mahar, A. A. Sahito, A. M. Soomro, G. S. Kaloi, "Dynamic Voltage Restorer—A comprehensive review", Energy Reports, Vol. 7, 2021, pp. 6786-6805.
- [12] S. F. Al-Gahtani et al. "A New Technique Implemented in Synchronous Reference Frame for DVR Control Under Severe Sag and Swell Conditions", IEEE Access, Vol. 10, 2022, pp. 25565-25579.
- [13] L. E. Christian, L. M. Putranto, S. P. Hadi, "Design of Microgrid with Distribution Static Synchronous Compensator (STATCOM) for Regulating the Voltage Fluctuation", Proceedings of the IEEE 7th International Conference on Smart Energy Grid Engineering, Oshawa, ON, Canada, 12-14 August 2019, pp. 48-52.
- [14] A. P.N.Tahim, D. J.Pagano, M. L. Heldwein, E. Ponce, "Control of interconnected power electronic converters in dc distribution systems", Proceedings of the XI Brazilian Power Electronics Conference, 2011, pp. 269-274.
- [15] N. Ghanbari, S. Bhattacharya, "Constant Power Load Challenges in Droop Controlled DC Microgrids", Proceedings of the 45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14-17 October 2019, pp. 3871-3876.
- [16] A. Engler, N. Sultanis, "Droop control in LV-grids", Proceedings of the International Conference on Future Power Systems, Amsterdam, Netherlands, 18 November 2005, pp. 1-6.

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