In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 12, No.2, April-June 2024, Serial Number 46

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

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

Indexed by:

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

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

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

Acknowledgement

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

(A-Z)

- Ahmadizad, Arman, University of Kurdistan, Sanandaj, Iran
- Boluki Speily, Omid Reza,Urmia University of Technology, Iran
- Badie, Kambiz, Tehran University, Iran
- Etezadifar, Pouria, Imam Hossein University (IHU), Tehran, Iran
- Fathi, Amir, Urmia University, Urmia, Iran
- Farsi, Hassan, University of Birjand, South Khorasan, Iran
- Farbeh, Hamed, Amirkabir University,Tehran, Iran
- Fadaeieslam, Mohammad Javad, Semnan University, Iran
- Ghasemzadeh, Ardalan, Urmia University of Technology, West Azerbaijan, Iran
- Jampoor, Mehdi, Quchan University of Technology, Razavi Khorasan, Iran
- Kasaei, Shohreh, Sharif University, Tehran, Iran
- Mirzaei, Abbas, Islamic Azad University, Ardabil, Iran
- Mohammadzadeh, Sajjad, University of Birjand, South Khorasan, Iran
- Munesh, Singh, Indian Institute of Information Technology Design & Manufacturing Jabalpur, India
- Minoofam, Seyed Amir Hadi, Qazvin Islamic Azad University, Qazvin, Iran
- Omid Mahdi, Ebadati, Kharazmi University, Tehran, Iran
- Pashazadeh, Saeid, Tabriz University, Tabriz, Iran
- Rasi, Habib, Shiraz University of Technology, Shiraz, Iran
- Rane, Milind, Vishwakarma Institute of Technology, Pune, Maharashtra, India
- Ramezani, Reza, Isfahan University, Isfahan, Iran
- Shamsi, Mahboubeh, University of Qom, Iran
- Saadatfar, Hamid, University of Birjand, Iran
- Shamsi, Mahboobeh, Qom university of technology, Iran
- Soleimanian Gharehchopogh, Farhad, Islamic Azad University Urmia, Iran
- Tourani, Mahdi, University of Birjand, South Khorasan, Iran
- Tanhaei, Mohammad, Ilam University, Ilam , Iran
- Zayyani, Hadi, Qom University of technology, Qom, Iran

Table of Contents

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 90-104
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

FLHB-AC: Federated Learning History-Based Access Control Using Deep Neural Networks in Healthcare System

Nasibeh Mohammadi[1], Afshin Rezakhani[2*], Seyd Hamid Haj Seydjavadi[3], Parvaneh Asghari[4]

[1]. Department of Computer Engineering, Islamic Azad University, Boroujerd Branch, Borujerd, Iran
[2]. Department of Computer Engineering, Faculty of Engineering, Ayatollah Boroujerdi University, Boroujerd, Iran
[3]. Department of Computer engineering, Shahed University, Tehran, Iran
[4]. Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

Received: 22 Oct 2023/ Revised: 04 Feb 2023/ Accepted: 03 Mar 2024

Abstract

Giving access permission based on histories of access is now one of the security needs in healthcare systems. However, current access control systems are unable to review all access histories online to provide access permission. As a result, this study first proposes a method to perform access control in healthcare systems in real time based on access histories and the decision of the suggested intelligent module. The data is used to train the intelligent module using the LSTM time series machine learning model. Medical data, on the other hand, cannot be obtained from separate systems and trained using different machine-learning models due to the sensitivity and privacy of medical records. As a result, the suggested solution employs the federated learning architecture, which remotely performs machine learning algorithms on healthcare systems and aggregates the knowledge gathered in the servers in the second phase. Based on the experiences of all healthcare systems, the servers communicate the learning aggregation back to the systems to control access to resources. The experimental results reveal that the accuracy of history-based access control in local healthcare systems before the application of the suggested method is lower than the accuracy of the access control in these systems after aggregating training with federated learning architecture.

Keywords: Healthcare System; History-Based Access Control; Intelligent Module; Deep Recurrent Networks; Federated Learning.

1- Introduction

A health information system (HIS) is a data management system for healthcare. This comprises systems for collecting, storing, managing, and transmitting a patient's electronic medical record (EMR); hospital operational management systems; and systems that support healthcare policy choices. Health information systems also include data management systems for healthcare practitioners and organizations. These technologies, for example, might be used to enhance patient outcomes, inform research, and impact policy and decision-making. Security is a critical concern in health information systems because they typically access, analyze, or store a substantial amount of sensitive data [1]. Access control refers to a set of procedures used to determine who or what has access to, uses, or changes what resources [2]. Access control is a critical topic in the design debates of physical security, information security, and network security to limit risks and threats. With the rapid advancement of computing and information technologies, classic access control models have become insufficient in terms of severe security needs, and new applications. ABAC models provide a more flexible way to deal with the authorization requirements of complex and dynamic systems. Modern access control systems and feature-based systems are becoming more popular [3,4]. Organizations are interested in adopting systems to regulate access to their resources that can best meet these demands as the use of access control systems expands in many industries such as IoT, cloud computing, and health care systems. During the COVID-19 time, for example, many businesses need a flexible approach to accessing software resources,

✉ **Afshin Rezakhani**
rezakhani@abru.ac.ir

allowing the user access to be adjusted based on the circumstances [5]. One of the existing challenges in access control systems is to use the previous accesses to grant the access permission at the next stage online and without interruption. Also, protecting patients' privacy while checking previous accesses is another challenge facing this article, which we try to solve it. The contributions of this paper are as follows:

- ✓ Using access histories in granting access permission to healthcare information systems.
- ✓ Using the deep recurrent network and intelligent module to provide online/real-time access requests.
- ✓ Maintaining the privacy of patient data in the HIS system and using local patient data to train the learning model
- ✓ Using federated learning to consolidate the training of local healthcare systems and improve the accuracy of the access control

The rest of this paper is as follows: Section 2 presents the background and fundamental concepts used in the paper background. In section 3, the surveys of related works are considered. The suggested approach and its components are explained in section 4. In section 5, the performance results and comparison with previous works are described. A discussion of our work is presented in section 6 and finally, section 7 concludes the paper.

2- Background

The basic ideas utilized in the article are defined in this section. The definitions of ABAC access control systems are explored first, followed by deep recurrent neural networks and federated learning model.

2-1- Attribute-based Access Control

The attribute-based access control (ABAC) is a type of access management model in which permission to perform a set of operations is determined by analyzing the attributes assigned to requesters, resources, and requested activities, as well as environmental conditions in some cases. Attributes are qualities that indicate a specific aspect of the requester and objects, environmental circumstances, or requested actions that the administrator has already established and assigned [4][6].

- ***Definition 1. Entities and Attributes:*** U, O, C, and OP are the system users (requesters), requested resources, defined particular conditions, and requested operations in the ABAC access control paradigm. Also, A_u, A_o, A_c, and A_{op} are the attributes of the requester, source, condition, and requested operation, respectively. Also, $E = U \cup O$, $\cup$ is the set of all entities, and $A = A_u \cup A_o \cup A_c \cup A_{op}$ is the set of all attributes of the entities mentioned above.
- ***Definition 2. Mapping Function:*** If $a \in A$ and $e \in E$, $F_{e,a}$ is the mapping function of the entity e on the attribute a. More precisely, the function $F_{e,a}(e,a)$ returns all the items that are related to the existence of e on attribute a. For example, $F_{a,e}$(Ali, location) = Tehran means that the location related to Ali is Tehran.
- ***Definition 3. Access request(req):*** An access request is a <u, o, c, op> where a requester u has requested operation op to access resource o in condition c. For example, <Ali, db1, Tehran, read> is an access request by Ali, who wants to read access to the source db1 from Tehran.
- ***Definition 4. Conventional access policy:*** Access permission is a sample(sample) sample=<req, g> decision g on request req_i in the access control system. sample can also be considered as access history and shows the access details.
- ***Definition 5. History-based access policy:*** If the access request is based on the records of sample <u, o, c, op>, the user u has requested the operation op to access the source o in the condition c where there is a c_j in c and it is essential to check the histories or check the prior accesses. For instance, *<db1, location.NewYork, $count_{100}$(access(db1, location(Tehran))<2), read, permit>* is an access policy of this type. If two conditions are met, the requester is granted access to db1. First, the requester's location must be in New York, and the number of accesses on db1 from the Tehran location should be no more than twice in the preceding 100 accesses. As a result, the access policy is established as <u, o, c, op, g> granting the requester g access in exchange for op access on o.
- ***Definition 6. Policy Repository(pr):*** A database containing all of the policies described in definitions 5 and 6. This repository's policies are organized into two broad groups. The first category includes only attributes related to the requester, resource, condition, and the requested operation (as seen in definition 1, we have designated the set of all these attributes with A), and the second category includes attributes that require checking access histories, which we denote by LA.
- ***Definition 7. History Access:*** If the current access request was made at time t, the access history provides a list of all accesses made from time t-k to time t-1. For example, if R_{t-k} access is provided at time t-k and R_{t-1} access is granted at time t-1,

the following access histories are defined at time t: $Hi(t)=\{R_{t-k}, R_{t-k+1}, \ldots, R_{t-1}\}$

2-2- Recurrent Neural Networks

Recurrent neural networks (RNNs) are artificial networks that are utilized in speech recognition, natural language processing, and sequential processing [7]. However, RNN features a feedback layer that feeds back the network's output as well as the next input. RNN can recall its previous input and use this information to process a sequence of inputs because of its internal memory. Simply said, RNNs incorporate a feedback loop that ensures that past knowledge is not lost and remains in the network. The following are the architecture types of RNNs.

2-2-1-Simple Recurrent Neural Network (RNN)

This is the most basic sort of recurrent network, yet it is still a viable alternative due to its modest number of parameters (when compared to GRU and LSTM networks) and reasonable accuracy in simple situations and short time series. The fundamental issue with simple RNNs is their limited memory, which results in vanishing and exploding gradients [7].

2-2-2-Long Short-Term Memory (LSTM)

LSTM networks are an upgraded version of RNNs that improves memory recall. In this sort of network, the problem of progressive fading of RNNs has been solved. LSTM is appropriate for time series categorization, processing, and prediction in the presence of time delays of unknown duration [8]. In addition, the LSTM network cell's inputs and outputs are as follows.

2-2-3-Gated Recurrent Units (GRUs)

The GRU recurrent neural network, like the LSTM, is intended to alleviate the RNN's short memory problem. Hidden layers are employed to handle and categorize input in the gated recurrent network rather than the state cell [9][10].

2-3- Federated Learning

The purpose of federated learning is to train a machine learning algorithm, i.e. deep neural networks, on multiple local datasets that exist at local nodes without explicitly exchanging data. The general principle consists of training local models on local data samples and exchanging parameters (e.g. weights and biases of the deep neural network) between these local nodes at some frequency to produce a global model shared between all nodes. The main difference between federated learning and distributed learning is in the assumptions made about the properties of the local datasets because the main goal of distributed learning is to parallelize the computing power whereas federated learning initiative aims to train heterogeneous datasets. While the goal of distributed learning is to train a single model on multiple servers, a common underlying assumption is that the local datasets are identically distributed and have approximately the same size. None of these hypotheses were made for federated learning. Instead, data sets are typically heterogeneous and their size may span several orders of magnitude. In addition, clients involved in federated training may be unreliable as they are subject to more failures or dropouts compared to distributed learning where nodes are typically data centers with powerful computing capabilities and are connected to fast networks [11] [12].

2-4- Healthcare System

A Healthcare platform enables doctors and their assistants to analyze data. Such an infrastructure should have things like simple equipment management, simple connections, data analysis, and intelligent data transformation [13]. Due to the vast amount of information, healthcare platforms must have the ability to accurately and timely analysis to provide the best analysis of various conditions. An overview of IoT-based healthcare is shown in Figure 1. The following components are essential in healthcare systems:

- ✓ Data collection using existing sensors
- ✓ Supporting a simple user interface for use by all patients and medical centers
- ✓ Access to infrastructure services and network services for all nodes in the network
- ✓ Increasing reliability, accuracy, durability, and strength in data storage and transmission.

Figure 1. An overview of a typical IoT-based healthcare system [14]

2-5- Related Works

This section will look at some recent studies in the field of the present subject. The authors of [15]

presented a blockchain-based access control system for GWAS with BFGF federated learning in their paper. Before training the local models, the framework uses automatic quality control (AQC) to assure the quality of the training data in this technique. It creates a blockchain authentication system to filter people. The authors of [16] focused on data security and privacy in industrial IoT systems utilizing machine learning models and federated learning models. They also investigated and contrasted other ways. A novel middleware for risk-based authorization and federated learning for health care (FRAMH) has been presented in [17], which provides risk-based access control for medical records. The authors employed a federated learning approach to assess health status risk and integrated it with blockchain to prevent unwanted access. Another study [18] offered "Hash and Signature-Based Policy-Based Encryption (hCP-ABES)," a cloud-based healthcare system for secure data storage and access control. The authors' proposed access control provides users with security, authentication, and secrecy when using medical data. Encryption and auditing procedures are employed to maintain the confidentiality and integrity of stored information. Access control mechanisms are usually used to govern data access during the data-sharing phase. In the data analysis process, machine learning algorithms are used to secure the privacy of massive medical data [19]. The authors of [20] suggested a framework to address issues such as information leaking in access granting. The authors suggested a federated learning framework for access control policies as well as a formal explanation of the policy transfer problem in attribute-based access control. Recent breakthroughs in the field of federated learning for cyber security and IoT security have been thoroughly addressed in [21]. This study's primary focus is on security, but it also explores different techniques for addressing FL-related performance difficulties that may compromise IoT security and performance. Another area of study is federated learning use in industrial systems [22]. This research looks into the FL prospects for next-generation networked industrial systems, as well as the problems of collaborative driving in connected and robotic autonomous vehicles. An approach for leveraging federated learning as a service with Decentralized Identities is proposed in [23]. The authors presented a DID-eFed system in which decentralized identities (DID) and a smart contract facilitate FL. DID offers flexible decentralized access management in the proposed system, and the smart contract provides a process with a few errors. [24] investigates FL in-depth, focusing on applications and operating systems, methods, and real-world applications. FL generates robust classifiers without requiring information disclosure, resulting in extremely secure privacy policies and access control rights. The authors in [25] propose a novel approach of combining CNN-LSTM with particle swarm optimization in the RBAC system. The convolutional neural network has extracted parsed SQL queries and long short-term memory was also suitable for modeling the temporal information of SQL queries. The paper [26] has considered an access control model for multi-channel heterogeneous networks based on deep reinforcement learning, referred to as multi-channel deep-reinforcement learning. To overcome the challenges of securing IoT devices, the authors in [27] have suggested a deep learning–based intrusion detection system to detect security vulnerabilities in IoT. The research [28] has protected the agreements dependent on ERC20 of controlled Ethereum-based Distributed Ledger Technology with cycles and capacities to get an all-surrounding framework for creating sure Cloud-Based Manufacturing jobs. effective attribute-based encryption is suggested in [29] which places part of the cryptography in the edge nodes as well as supports attribute updates and flexible control. To address the problem of scalability in access control, the authors in [30] have proposed an enhanced Bell–LaPadula model and categorized the peers and transactions in different clearance and security levels. In the article [31], the authors have proposed a blockchain-based approach that provides a decentralized EHR and smart-contract-based service automation without compromising the system's security and privacy. The paper [32] discussed some gaps within the existing access control strategies for health care. To fill this gap, the authors have proposed a secure access control model to control access in the healthcare system. Their solution has used the location of the user for providing secure access control views in the healthcare system. The article [33] has planned to create a novel solution based on blockchain technology that locates the patient in charge of granting and revoking access permissions for healthcare enterprises and providers to meet privacy regulations. Motivated by the research gaps, the paper [34] proposed a scheme, that integrates a blockchain (BC)-based confidentiality-privacy (CP) preserving scheme. In the paper [35], the authors have discussed how leveraging blockchain for healthcare data control can lead to better improvements. they presented the key blockchain features and characteristics. The paper [36] has focused on privacy issues in smart context-aware healthcare within the Electronic Transfer of Prescriptions. The access control models may expose user privacy to an attacker. To tackle this problem,

the authors [37] have used the cuckoo filter to disguise the right of entry policy to safeguard the private information of the owner. The inference assault has affected medical records. The paper [38] has proposed a new blockchain-based lightweight access control model. The scheme has used blockchain to create a trusted network by a special mechanism. The paper [39] reviewed the recent trends and critical requirements for blockchain-based and IoT access management. The authors showed several important views of blockchain, including decentralized control, secure storage, and sharing information for IoT access control. Deep learning and artificial intelligence frameworks are introduced in the paper [40] to improve cyber security such as access control. The authors in [41] have proposed a novel model by implementing a specific cryptography algorithm in which they used the Key generation scheme of RSA to encrypt health data. The paper [42] has proposed a deep learning-based anomaly detection method composed of estimation and classification models applied to a subdomain in healthcare systems. the authors [43] have conducted a universal review of federated learning systems. To get a clear flow and guide future probes, they introduced the definition of federated learning systems and analyzed the system sections. The authors of [44] presented a method for automatically learning ABAC policy rules from the access logs of a system. In the proposed approach, an unsupervised learning-based algorithm is used to recognize patterns in extracting ABAC authorization rules. In [45], an efficient and simple method has been proposed to verify the access control policy using a machine learning classification algorithm. Cotrini et al. [46] suggested an approach for deriving some rules from randomly distributed histories. In [49], a novel method was suggested for secured and integrated access control in the SIEM. The key points where the SIEM accesses the information within the software were specified and policies for access control were developed. To achieve energy efficiency in the network some simplification strategies have to be carried out not only in the Medium Access Control (MAC) layer but also in the network and transport layers [50].

3- Problem Definition

Although giving access authorization in HIS systems necessitates the use of feature-based access control, only limited features such as the requester's location and time have been used to validate access permission thus far. In the first step of this research, "access history" is proposed as an essential component in HIS system access control. A pulse based on a log, for example, is described as "a certain drug is administered to a patient if it is prescribed by at least two doctors." For such a pulse, the prior logs must be examined. To study access control policies based on log conditions, LSTM and GRU deep learning models are utilized, which must be triggered live during access requests. Deep learning is employed since you cannot spend a lot of time verifying access histories while seeking access. As a result, memory neural networks such as LSTM and GRU will be useful.

The limited number of samples available to train the learning model in local HIS systems presented a challenge during the initial step. Furthermore, because of privacy concerns, it is not viable to gather all medical data in HIS systems and train the intelligent learning model (previous step). As a result, the federated learning architecture is used in the second stage of the proposed approach, and instead of transferring data to the server, the suggested model is produced in the server and delivered to the HIS systems. After training the model in HIS systems, the model's output is forwarded to the server, which aggregates the results.

3-1- Challenges and Requirements

To develop log-based access control that can be applied to a wide range of real-world scenarios, we identify the following challenges and requirements:

- Online access request: Log-based resource access requests are submitted online and require a prompt, no-delay response with a short time order. However, the problem of the access control system based on access histories necessitates a review of past logs as well as the processing and time loads. As a result, the proposed model should be able to handle a huge number of online queries.
- Correctness in log-based access control: Because the extracted access decision must result in the same access decision as the policy repository, the suggested log-based access control system's prediction (response) must be compatible with the result of the original permission in the policies. An inconsistent answer may arise in instances where previously approved access is refused (more restrictive) or the system allows unlawful access (less restrictive).
- The complexity of log-based policies: The complexity of log-based policies is one of the most important challenges in access control systems. In many policies, it is not required to check logs, but in some policies, logs should be

checked. Such conditions will cause complexity in meeting all the conditions with a reciprocal effect on each other.

- Privacy Violation: Considering that medical records in HIS systems contain private information of patients, it is important to protect privacy during research and data review.

3-2- Evaluation Metrics

How well these accesses match the original accesses is one of the metrics for evaluating the quality of the accesses provided in the proposed approach (given by the original pulses). In other words, the proposed method's access result is compared to the original pulses' access result, and the quality of the suggested access control system is evaluated. For example, if the suggested method's prediction for an access request is to grant permission and access permission is granted in the main policy, the quality of the proposed approach will improve. The following definitions are taken into account for a more complete study of the proposed approach's evaluation metrics.

- ***Definition 8 (TP definition):*** If an access request is req_i=< $attr_s$, $attr_o$, opr, act > and the proposed approach prediction for req_i is equal to $pred_i$, and also if the original access policy is rul_j=< $attr_s$, $attr_o$, opr, act> and the label (access) of the original policy equivalent to the request is equal to lab_j, then TP is defined as follows.
  $TP_{<req,pred>}$ = | <r> ϵ DS | (lab(rul(r)) == permit) && (pred(req(r)) == permit) |
- ***Definition 9 (Definition of TN):*** Despite the assumptions of Definition 8, TN is defined as follows. This means that the prediction made by the proposed approach with the original policy label corresponding to that request is equal to *denial*.
  $TN_{<req,\ pred>}$ = | <r> ϵ DS | (lab(rul(r)) == deny) && (pred(req(r)) == deny) |
- ***Definition 10 (Definition of FP):*** Despite the above assumptions, FP is defined as follows. This means that the prediction made by the proposed approach is *permitted*, but the label of that request in the original policy is equal to deny, and it shows the wrong prediction in the proposed approach.
  $FP_{<req,\ pred>}$ = | <r> ϵ DS | (lab(rul(r)) == deny) && (pred(req(r)) == permit) |
- ***Definition 11 (Definition of FN):*** If the prediction made by the proposed approach is *denied* and the label of that request in the original policy is equal to permit, it means a wrong prediction in the proposed approach, which is defined as follows.
  $FN_{<req,\ pred>}$ = | <r> ϵ DS | (lab(rul(r)) == permit) && (pred(req(r)) == deny) |
- ***Definition 12:*** by calculating the above metrics, Accuracy = $\frac{TP + TN}{TP + FN + TN + FP}$ can be defined to determine the accuracy of the proposed approach.

Table 1 presents the symbols used in this article.

Table 1. Notations

Notation	***Definition***
U, O, C, OP	Users(requesters), resources, conditions, and operations
A_u, A_o, A_c, A_{op}	attributes of the requester, source, condition, operation
E, A	*U* ∪ *O*, set of all attributes of the entities
a, *e*, c_j	a∈A, e∈E, c_i∈c
G	Access permission
$F_{e,a}$	Mapping function of the entity e on the attribute a
C	Conditions in access histories
t, k	Time
R_i	Access in time i
Hi	access histories in time t
m_i	Equivalent method with any condition C
M	set of methods {m1, m2, ...}
F#, S#	The number of data features, The number of data samples
req_i	Access request i
Q_i	Set of features reqi
Train_Set	Training dataset in conditional dataset
NLA	Set {nla1, nla2, …}
LA	Set {la1, la2, …}
pr	Policy Repository
PAP	Policy administration point
PEP	Policy enforcement point
PDP	Policy decision point
PIP	Policy information point
DT, KNN, NN, SVM	decision tree, k-nearest neighbor, neural network, Support vector machine
$a^{<t>}$, $c^{<t>}$	Current stage status
$a^{<t-1>}$, $c^{<t-1>}$	The status of the previous stage
$x^{<t>}$,$\hat{y}^{<t>}$	Input, output (prediction) of the recurrent network
w_a, W_o, W_f, W_u, W_y, w_c	Weight vectors
$_t\hat{c}$	status of the new stage
σ	sigmoeid activity function
$_{u\Gamma}$, $_{f\Gamma}$	Update gate, Forget Gate
b_y, b_u,b_c, b_a	Bias
g_1,g_2	tanh activity functions

4- Intelligent FLACH Approach

This study is primarily concerned with providing access permission in healthcare systems based on the aggregation of local healthcare systems' knowledge and experiences. Because medical records are sensitive and private, it is not practical to collect data from various healthcare systems and train them using machine learning models. The federated learning approach is proposed in this paper to employ machine learning algorithms remotely on distinct local systems and aggregate the knowledge gained in

the servers. The servers re-send the aggregation of various learnings to the clients to regulate resource access based on the experiences of all local healthcare systems. Figure 2 depicts the general architecture of the suggested approach, which we call FLACH.

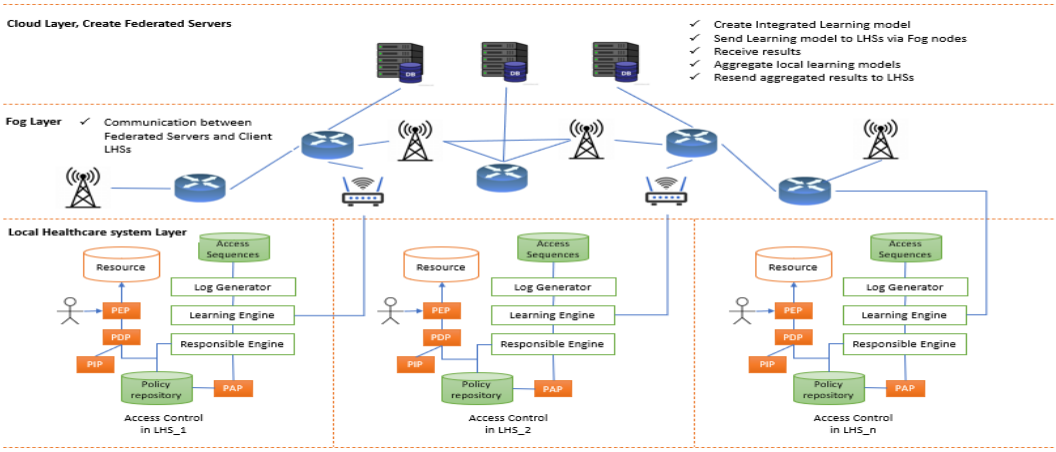


Figure2. Proposed multi-layer architecture

Before delving into the FL architecture, some key assumptions must be addressed. 1) Using access histories to authorize access is one of the requirements of healthcare systems. A nurse, for example, can inject a specific drug into a patient if it has been ordered by two general practitioners or a specialist. 2) Deep learning models based on time series, such as Simple RNN, LSTM, and GRU, are appropriate for granting access based on previous access histories. 3) Learning models like LSTM and GRU employ the order of subsequent accesses to grant access authorization. 4) The privacy of medical data is a distinguishing aspect of healthcare systems. As a result, patient information cannot be withdrawn from the local healthcare system.

The following sections describe the various components of the suggested method (which are explained in depth in the next section)

1. Federated learning architecture for LHS knowledge aggregation: The data is trained locally in each LHS's access control system before being utilized to give access. However, the difficulty with this proposal is the lack of limited and sensitive data, resulting in inadequate training for the proposed intelligent module. As a result, the technique of federated learning architecture is presented to aggregate the knowledge of intelligent access control systems (which exist in local healthcare systems).
2. An intelligent module within the access control system: to train the access control system, deep learning methods based on time series, such as LSTM or GRU, are used. When the policy repository incorporates access control rules based on checking access histories, the intelligent module in the access control system is required and unavoidable. As a result, an intelligent module for each local healthcare system is installed alongside the access control system.
3. Customized access control: When intelligent modules are installed in local access control systems and the knowledge of various modules is combined, the intelligent access control system is customized.

4-1- Design of Federated Learning Layers

First, servers are created in the cloud layer to aggregate training in local healthcare systems. These servers are cloud-based and have no access to medical records (data) in healthcare systems. These servers' primary function is to build a machine learning model based on time series and then distribute it to all local systems.

First, using Algorithm 1, a machine learning model based on time series is generated in the cloud layer servers. This article focuses on the use of LSTM and GRU models. In this algorithm, the learning model is defined first, followed by the necessary pre-processing. The necessary layers are then inserted, and the model compiles and begins training.

Algoithm1. Machine learning model created in server in the smart machine

Procedure *Create-Model ()*
Input: *Conditions, dataset*
Output:
Forever ()
{
model ← Sequential ()
dataset ← Preprocessing(dataset)
model ← CreateModel ()
model. Add (layers, optimizers, activations, …)
…
model.add (Dense, Dropout, …)
model. Compile (Train_Set)
model. Fit (Train_Set)
}

The model developed on the server is transferred via fog layer nodes to the intelligent module of local healthcare systems. There is no data transfer in this transfer; just the learning model is sent to the fog nodes to be sent to the local LHS systems .

After training the models in the intelligent module of the local systems, the results are transmitted back to the servers to be aggregated and re-distributed to the local systems via the fog layer nodes. Figure 3 depicts the sequence of generating the model on the server, delivering it to the LHSs, and resending the trained weights from the LHSs to the servers via the fog layer nodes.

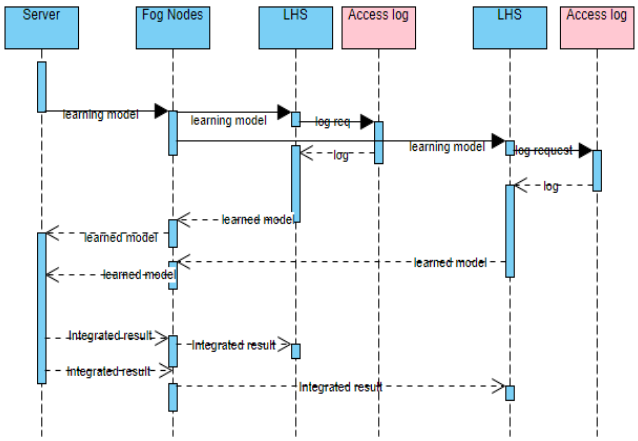


Figure 3. The sequence number of training models transfers from the server to the local healthcare systems

4-2- Intelligent Module Design

As previously stated, client layer design is a component of federated learning architecture. The proposed intelligent module is the federated architecture's client component. As a result, at this stage, the intelligent module is generated locally within the healthcare system's access control system. The learning engine, reaction engine, and access logs are all part of the local smart module. Machine learning models based on time series are assembled and run in the learning engine. It trains the learning model in the engine using local access records. In addition, the response engine will be employed in the access control system to answer access requests. When the policy repository contains policies that need access histories to be reviewed, the response engine is used (rather than checking all previous accesses). Figure 4 depicts the proposed module's components and their connections.

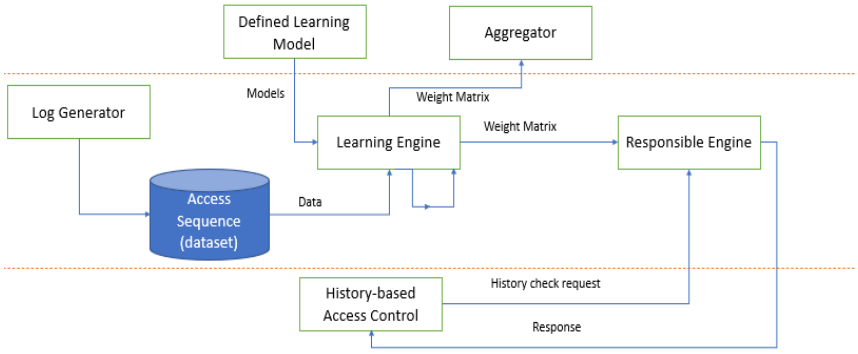


Figure 4. Components and connections of the intelligent module within the access control system

4-3- History-based Access Control Design

The proposed access control system uses an attribute-based model (access history). According to this model, the requestor sends the request to the policy enforcement point or PEP. This component sends the request to the PDP. PDP makes decisions through two main units policy repository and PIP. The PIP informs the current status of the requester and the resource. Also, all policies are in the policy repository.

So initially, the PIP delivers all the requester and resource attributes to the PDP. The PDP then retrieves access control policies associated with the requester from the policy repository. If the history corresponding to the requester in the policy repository does not contain log-based properties, the access permission is granted directly from the repository and there is no need to call the intelligent module's response engine. However, if the requester's history in the repository contains log-based properties, the PDP refers to the response engine, and the response of this engine is issued as a result of the access permission. Algorithm 2 shows the implementation of the access control system, which can implement both the current ABAC systems and the systems based on access histories.

Algoithm2. Access control algorithm

Procedure *AccessControl()*
Input: $request_i$
Output: *grant*
    $req_i \leftarrow request_i$
    PEP sends req_i *to PDP*
    PDP asks information of req_i *from PIP*
    $Q_i \leftarrow PIP\ (req_i)$
    For all $\boldsymbol{pr}_k \in \boldsymbol{policyRepository}$
        If $\boldsymbol{pr}_k$ *[NLA-LA] ==* Q_i
            If $\boldsymbol{pr}_k$ *[LA] ==* $\emptyset$
                Return $\boldsymbol{pr}_k$ *[label]*
            Else if pr_k *[LA] !=* $\emptyset$
                Return (SmartMachine. Result [req_i*])*
}

5- Experimental Results

Based on access histories, a prototype of an access control system is created. The evaluation conditions and datasets utilized are explained first in this section. The evaluation parameters are then set, and the performance of the proposed access control system in the federated learning framework is carefully examined. Finally, the proposed method's performance is compared with previous studies.

5-1- Evaluation Conditions

A system with an Intel Core i7 processor and 16 GB RAM is utilized to analyze the suggested model. Anaconda and Python are used to transfer time series-based learning models from servers to clients and vice versa. We conducted our studies on six datasets and twelve situations. How well the accesses match the original access is one of the major metrics for evaluating access accuracy. In other words, the proposed system's access results are compared to the key policies' access results, and the suggested access control system's quality is evaluated. For example, if the proposed system predicts that an access request will be granted and access authorization is granted in the primary policy, the quality of the suggested system would improve.

5-2- Datasets

We employ six datasets, including real and conditional datasets, to evaluate the performance of the proposed approach. Real datasets such as Kaggle and Amazon UCI datasets [47][48] are employed, as well as conditional datasets. The RESOURCE, MGR ID, ROLE ROLLUP1, ROLE ROLLUP2, ROLE DEPTNAME, ROLE TITLE, ROLE FAMILY DESC, ROLE FAMILY, ROLE CODE, and a label field are among the ten features in the Kaggle dataset.

In addition, the UCI dataset has five features and approximately 716K samples, including ACTION, TARGET NAME, LOGIN, REQUEST DATE, AUTHORIZATION DATE, and labels.
Four conditional datasets are also employed, where each one is generated based on the criteria and sequence of accesses. These four datasets are created using Python and based on the description of numerous requirements. They comprise features such as Name, Role, Time, Location, Sensitivity, and labels. Table 2 shows the general characteristics of the datasets used. D is the name of the dataset used in this table, S# is the number of dataset samples, F# is the number of dataset features, P^+ is the number of samples labeled 1 and P^- is the number of samples labeled zero.

Table 2. Characteristics of the employed datasets

#	D	S#	F#	P^+	p^-
d_K	*Amazon Kaggle*	*32769*	*10*	*30872*	*1897*
$d_{K,1}$ - $d_{K,10}$	*Datasets $d_{K,1}$ - $d_{K,10}$ are created from dataset d_K*	*32769*	*10*	*30872*	*1897*
d_U	*Amazon UCI*	*716063*	*3*	*705152*	*10911*
$d_{U,1} - d_{U,10}$	*Datasets $d_{U,1}$ - $d_{U,10}$ are created from dataset d_U*	*716063*	*3*	*705152*	*10911*
d_{C1}	*Conditional Dataset1*	*10000*	*6*	*9105*	*895*
$d_{C1,1} - d_{C1,10}$	*Datasets $d_{C1,1} - d_{C1,10}$ are created from dataset d_{C1}*	*10000*	*6*	*9105*	*895*
d_{C2}	*Conditional Dataset2*	*10000*	*6*	*7022*	*2978*
$d_{C2,1} - d_{C2,10}$	*Datasets $d_{C3,1} - d_{C3,10}$ are created from dataset d_{C3}*	*10000*	*6*	*7022*	*2978*
d_{C3}	*Conditional Dataset2*	*10000*	*6*	*7022*	*2978*
$d_{C3,1} - d_{C3,10}$	*Datasets $d_{C3,1} - d_{C3,10}$ are created from dataset d_{C3}*	*50000*	*6*	*5888*	*4112*
d_{C4}	*Conditional Dataset4*	*10000*	*6*	*29833*	*20167*
$d_{C4,1} - d_{C4,10}$	*Datasets $d_{C4,1} - d_{C4,10}$ are created from dataset d_{C4}*	*10000*	*6*	*29833*	*20167*

5-3- Results

In the first stage, the performance of the proposed method is investigated; conventional classification models such as KNN, SVM, DT, and NN, as well as time series classification models such as Simple RNN, LSTM, and GRU, have been implemented on datasets d_K, d_U, d_{C1}, d_{C2}, d_{C3}, and d_{C4}, and the results are shown in figures 5 and 6. As can be observed, models based on time series, such as LSTM, outperform conventional classification methods in terms of correct identification of the access control system.

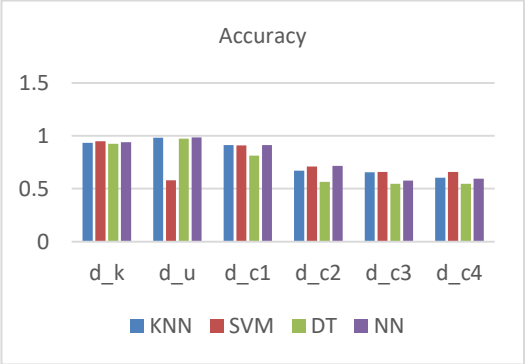


Figure 5. Accuracy of the proposed approach in algorithms without time series in different datasets

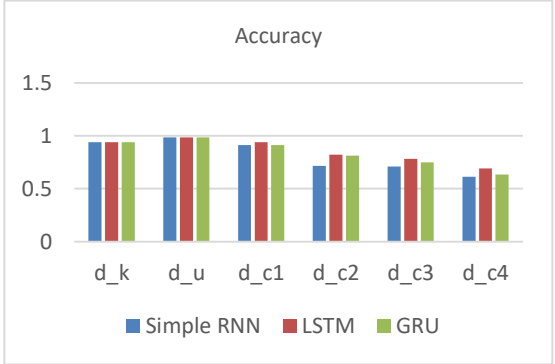


Figure 6. Accuracy of the proposed approach in time series algorithms in different datasets

The performance depicted in the figures above is estimated assuming that all data is accessible. However, the suggested method places the data in local healthcare systems and prevents it from being forwarded to servers. As a result, the dataset d_K is partitioned into ten smaller datasets numbered d_{K1}

through d_{K10}. Similarly, $d_{U,1}$ to $d_{U,10}$, $d_{C1,1}$ to $d_{C1,10}$, $d_{C2,1}$ to $d_{C2,10}$, $d_{C3,1}$ to $d_{C3,10}$, and $d_{C4,1}$ to $d_{C4,10}$ are formed from d_U datasets, as are d_{C1}, d_{C2}, d_{C3}, and d_{C4}. The separated datasets are located in local healthcare systems (Client), and none of these data are stored on the server owing to privacy concerns. On the server, a time series-based learning model is defined. This model is then given to ten clients depending on the written codes. The received model is applied to the data in clients, and the model is trained. The models' output (weight matrix) is then provided to the servers. The results are aggregated on the servers and resent to the clients. Table 3 displays the acquired results.

Table 3. Comparing the accuracy of the trained model in Local HIS systems and aggregated via federated learning architecture

	$partial_1$ DS	$Partial_2$ DS	$Partial_3$ DS	$Partial_4$ DS	$Partial_5$ DS	$Partial_6$ DS	$Partial_7$ DS	$Partial_8$ DS	$Partial_9$ DS	$partial_{10}$ DS	Total DS
d_K	ds=$d_{K,1}$ Acc=0.931	ds=$d_{K,2}$ Acc=0.925	ds=$d_{K,3}$ Acc=0.913	ds=$d_{K,4}$ Acc=0.933	ds=$d_{K,5}$ Acc=0.928	ds=$d_{K,6}$ Acc=0.920	ds=$d_{K,7}$ Acc=0.919	ds=$d_{K,8}$ Acc=0.926	ds=$d_{K,9}$ Acc=0.921	ds=$d_{K,10}$ Acc=0.918	Acc=0.939
d_U	ds=$d_{U,1}$ Acc=0.961	ds=$d_{U,2}$ Acc=0.971	ds=$d_{U,3}$ Acc=0.972	ds=$d_{U,4}$ Acc=0.975	ds=$d_{U,5}$ Acc=0.973	ds=$d_{U,6}$ Acc=0.981	ds=$d_{U,7}$ Acc=0.972	ds=$d_{U,8}$ Acc=0.969	ds=$d_{U,9}$ Acc=0.980	ds=$d_{U,10}$ Acc=0.970	Acc=0.985
d_{C1}	ds=$d_{C1,1}$ Acc=0.913	ds=$d_{C1,2}$ Acc=0.921	ds=$d_{C1,3}$ Acc=0.911	ds=$d_{C1,4}$ Acc=0.933	ds=$d_{C1,5}$ Acc=0.925	ds=$d_{C1,6}$ Acc=0.931	ds=$d_{C1,7}$ Acc=0.937	ds=$d_{C1,8}$ Acc=0.929	ds=$d_{C1,9}$ Acc=0.934	ds=$d_{C1,10}$ Acc=0.928	Acc=0.941
d_{C2}	ds=$d_{C2,1}$ Acc=0.811	ds=$d_{C2,2}$ Acc=0.814	ds=$d_{C2,3}$ Acc=0.797	ds=$d_{C2,4}$ Acc=0.921	ds=$d_{C2,5}$ Acc=0.812	ds=$d_{C2,6}$ Acc=0.792	ds=$d_{C2,7}$ Acc=0.819	ds=$d_{C2,8}$ Acc=0.824	ds=$d_{C2,9}$ Acc=0.810	ds=$d_{C2,10}$ Acc=0.799	Acc=0.823
d_{C3}	ds=$d_{C3,1}$ Acc=0.751	ds=$d_{C3,2}$ Acc=0.743	ds=$d_{C3,3}$ Acc=0.766	ds=$d_{C3,4}$ Acc=0.759	ds=$d_{C3,5}$ Acc=0.739	ds=$d_{C3,6}$ Acc=0.752	ds=$d_{C3,7}$ Acc=0.750	ds=$d_{C3,8}$ Acc=0.744	ds=$d_{C3,9}$ Acc=0.753	ds=$d_{C3,10}$ Acc=0.753	Acc=0.783
d_{C4}	ds=$d_{C4,1}$ Acc=0.622	ds=$d_{C4,2}$ Acc=0.634	ds=$d_{C4,3}$ Acc=0.651	ds=$d_{C4,4}$ Acc=0.957	ds=$d_{C4,5}$ Acc=0.641	ds=$d_{C4,6}$ Acc=0.648	ds=$d_{C4,7}$ Acc=0.636	ds=$d_{C4,8}$ Acc=0.659	ds=$d_{C4,9}$ Acc=0.645	ds=$d_{C4,10}$ Acc=0.660	Acc=0.691

5-4- Comparison with Recent Studies

To demonstrate the outperformance of the proposed approach, its performance is compared with that of Karimi [44] and Cotrini [46]. These methods have calculated the accuracy of their methods using machine learning algorithms. The results show the higher performance of the proposed method compared with others in problems that include time series conditions. Employing the machine learning models based on time series in d_K, d_U, d_{C1}, d_{C2}, d_{C3}, and d_{C4} using federated learning has a higher accuracy compared to the method presented by Karimi and Cotrini, as shown in Figure 7. The accuracy of the suggested model, which is based on time series, is compared with the research of Karimi and Cotrini in each local data corresponding to d_{K1} to d_{K10}, as well as d_{U1} to d_{U10}, $d_{C1,1}$ to $d_{C1,10}$, $d_{C2,1}$ to $d_{C2,10}$, and $d_{C3,1}$ to $d_{C3,10}$, $d_{C4,1}$ to $d_{C4,10}$ and shown in Table 4.

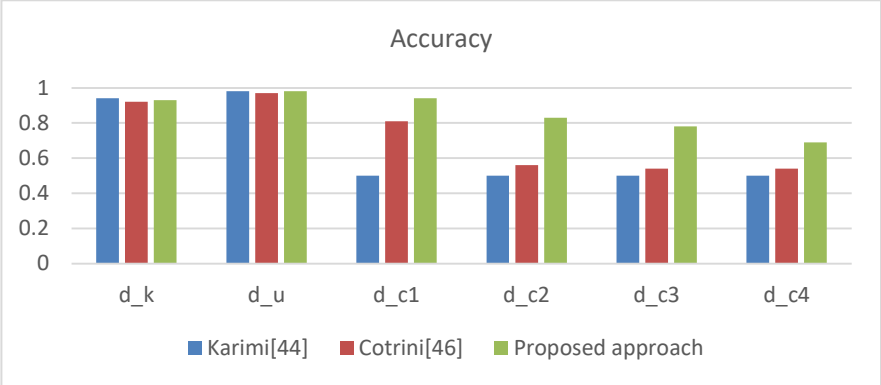


Figure 7. The accuracy comparison of the proposed method with other methods

Table 4. Comparing the accuracy of the trained model in Local HIS systems in each local data

	$partial_1$ DS	$Partial_2$ DS	$Partial_3$ DS	$Partial_4$ DS	$Partial_5$ DS	$Partial_6$ DS	$Partial_7$ DS	$Partial_8$ DS	$Partial_9$ DS	$partial_{10}$ DS	Aggregated DS
d_K *Karimi[44] Cotrini[46] Proposed approach*	$ds=d_{K,1}$ Acc=0.934 Acc=0.845 Acc=0.931	$ds=d_{K,2}$ Acc=0.931 Acc=0.844 Acc=0.925	$ds=d_{K,3}$ Acc=0.93 Acc=0.841 Acc=0.913	$ds=d_{K,4}$ Acc=0.943 Acc=0.829 Acc=0.933	$ds=d_{K,5}$ Acc=0.948 Acc=0.839 Acc=0.928	$ds=d_{K,6}$ Acc=0.937 Acc=0.839 Acc=0.920	$ds=d_{K,7}$ Acc=0.939 Acc=0.840 Acc=0.919	$ds=d_{K,8}$ Acc=0.946 Acc=0.829 Acc=0.936	$ds=d_{K,9}$ Acc=0.941 Acc=0.824 Acc=0.921	$ds=d_{K,10}$ Acc=0.938 Acc=0.834 Acc=0.918	Acc=0.943 Acc=0.849 Acc=0.939
d_U *Karimi[44] Cotrini[46] Proposed approach*	$ds=d_{U,1}$ Acc=0.961 Acc=0.673 Acc=0.961	$ds=d_{U,2}$ Acc=0.971 Acc=0.687 Acc=0.971	$ds=d_{U,3}$ Acc=0.965 Acc=0.671 Acc=0.972	$ds=d_{U,4}$ Acc=0.970 Acc=0.693 Acc=0.975	$ds=d_{U,5}$ Acc=0.975 Acc=0.682 Acc=0.973	$ds=d_{U,6}$ Acc=0.984 Acc=0.679 Acc=0.981	$ds=d_{U,7}$ Acc=0.972 Acc=0.691 Acc=0.972	$ds=d_{U,8}$ Acc=0.979 Acc=0.688 Acc=0.965	$ds=d_{U,9}$ Acc=0.976 Acc=0.682 Acc=0.980	$ds=d_{U,10}$ Acc=0.965 Acc=0.689 Acc=0.970	Acc=0.984 Acc=0.704 Acc=0.982
d_{C1} *Karimi[44] Cotrini[46] Proposed approach*	$ds=d_{C1,1}$ Acc=0.551 Acc=0.480 Acc=0.913	$ds=d_{C1,2}$ Acc=0.524 Acc=0.488 Acc=0.921	$ds=d_{C1,3}$ Acc=0.537 Acc=0.472 Acc=0.911	$ds=d_{C1,4}$ Acc=0.542 Acc=0.485 Acc=0.933	$ds=d_{C1,5}$ Acc=0518 Acc=0.49 Acc=0.925	$ds=d_{C1,6}$ Acc=0.543 Acc=0.491 Acc=0.931	$ds=d_{C1,7}$ Acc=0.517 Acc=0.47 Acc=0.937	$ds=d_{C1,8}$ Acc=0.498 Acc=0.492 Acc=0.929	$ds=d_{C1,9}$ Acc=0.519 Acc=0.486 Acc=0.934	$ds=d_{C1,10}$ Acc=0.523 Acc=0.487 Acc=0.928	Acc=0.556 Acc=0.501 Acc=0.941
d_{C2} *Karimi[44] Cotrini[46] Proposed approach*	$ds=d_{C2,1}$ Acc=0.475 Acc=0.459 Acc=0.811	$ds=d_{C2,2}$ Acc=0.490 Acc=0.461 Acc=0.814	$ds=d_{C2,3}$ Acc=0.494 Acc=0.469 Acc=0.797	$ds=d_{C2,4}$ Acc=0.489 Acc=0.471Acc=0.921	$ds=d_{C2,5}$ Acc=0.472 Acc=0.472 Acc=0.812	$ds=d_{C2,6}$ Acc=0.499 Acc=0.457 Acc=0.792	$ds=d_{C2,7}$ Acc=0.491 Acc=0.479 Acc=0.819	$ds=d_{C2,8}$ Acc=0.482 Acc=0.466 Acc=0.824	$ds=d_{C2,9}$ Acc=0.484 Acc=0.463 Acc=0.810	$ds=d_{C2,10}$ Acc=0.497 Acc=0.478 Acc=0.799	Acc=0.521 Acc=0.482 Acc=0.823
d_{C3} *Karimi[44] Cotrini[46] Proposed approach*	$ds=d_{C3,1}$ Acc=0.489 Acc=0.392 Acc=0.751	$ds=d_{C3,2}$ Acc=0.494 Acc=0.387 Acc=0.743	$ds=d_{C3,3}$ Acc=0.487 Acc=0.399 Acc=0.766	$ds=d_{C3,4}$ Acc=0.492 Acc=0.404 Acc=0.759	$ds=d_{C3,5}$ Acc=0492 Acc=0.385 Acc=0.739	$ds=d_{C3,6}$ Acc=0.501 Acc=0.390 Acc=0.752	$ds=d_{C3,7}$ Acc=0.497 Acc=0.399 Acc=0.750	$ds=d_{C3,8}$ Acc=0.498 Acc=0.385 Acc=0.744	$ds=d_{C3,9}$ Acc=0.489 Acc=0.409 Acc=0.753	$ds=d_{C3,10}$ Acc=0.493 Acc=0.396 Acc=0.753	Acc=0.503 Acc=0.413 Acc=0.783

d_{C2}	ds=$d_{C2,1}$	ds=$d_{C2,2}$	ds=$d_{C2,3}$	ds=$d_{C2,4}$	ds=$d_{C2,5}$	ds=$d_{C2,6}$	ds=$d_{C2,7}$	ds=$d_{C2,8}$	ds=$d_{C2,9}$	ds=$d_{C2,10}$	
Karimi[44]	Acc=0.466	Acc=0.580	Acc=0.457	Acc=0.455	Acc=0.460	Acc=0.471	Acc=0.484	Acc=0.472	Acc=0.472	Acc=0.457	Acc=0.489
Cotrini[46]	Acc=0.377	Acc=0.386	Acc=0.392	Acc=0.389	Acc=0.374	Acc=0.367	Acc=0.390	Acc=0.387	Acc=0.375	Acc=0.391	Acc=0.401
Proposed approach	Acc=0.622	Acc=0.634	Acc=0.651	Acc=0.957	Acc=0.641	Acc=0.648	Acc=0.636	Acc=0.659	Acc=0.645	Acc=0.660	Acc=0.691

6- Discussion

The proposed approach's performance has been carefully investigated in six datasets, and it can be observed that the accuracy of the proposed access control works well in d_K and d_U datasets using traditional machine learning methods such as SVM, but in conditional datasets, regardless of the circumstances, because access is complicated depending on logs, the techniques outlined above have poorer accuracy in granting right permission. Meanwhile, with more complex conditions in d_{C1} to d_{C4} datasets, the suggested method employing deep recurrent networks, particularly the LSTM network, outperforms other algorithms. Another advantage of the proposed method is that it provides internet access in a short time, which is due to the use of hidden memory in the recurrent neural network, as well as the training of this network during each access and testing the network at the time of access request so that no additional burden is imposed on the access control system when a new request is received. Furthermore, using federated learning architecture protects user privacy in HIS systems and integrates knowledge from other systems to allow access to users.

7- Conclusion

In this paper, a new approach for granting access authorization in healthcare systems online based on intelligent module decisions was provided. The utilized module responds to access requests using time series learning models such as LSTM and GRU. Local data from healthcare systems cannot also be aggregated in a single dataset. As a result, federated learning architecture and machine learning algorithms were remotely applied to various healthcare systems. The servers were in charge of gathering various learnings and relaying them to the local systems to regulate access based on the experiences of all systems. The experimental results reveal that the performance of the access control system in local systems after implementing the federated learning architecture and aggregating the knowledge of local systems is lower than the performance of this system before implementing the proposed approach.

References

[1] Haux, R. Health information systems–past, present, future. International journal of medical informatics, 75(3-4), 268-281, 2006.

[2] Ravidas, S., Lekidis, A., Paci, F., & Zannone, N. Access control in Internet-of-Things: A survey. Journal of Network and Computer Applications, 144, 79-101, 2019.

[3] Ding, S., Cao, J., Li, C., Fan, K., & Li, H. A novel attribute-based access control scheme using blockchain for IoT. IEEE Access, 7, 38431-38441, 2019.

[4] Hu, V. C., Kuhn, D. R., Ferraiolo, D. F., & Voas, J. Attribute-based access control. Computer, 48(2), 85-88, 2015.

[5] Wouters, O. J., Shadlen, K. C., Salcher-Konrad, M., Pollard, A. J., Larson, H. J., Teerawattananon, Y., & Jit, M. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. The Lancet, 397(10278), 1023-1034, 2021.

[6] Hu, V. C., Ferraiolo, D., Kuhn, R., Friedman, A. R., Lang, A. J., Cogdell, M. M., ... & Scarfone, K. Guide to attribute-based access control (abac) definition and considerations (draft). NIST special publication, 800(162), 1-54, 2013.

[7] Zaremba, W., Sutskever, I., & Vinyals, O. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329, 2014.

[8] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306, 2020.

[9] Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078, 2017.

[10] Lipton, Z. C., Berkowitz, J., & Elkan, C. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019, 2015.

[11] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. A survey on federated learning. Knowledge-Based Systems, 216, 106775, 2021.

[12] Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access, 8, 140699-140725, 2020.

[13] Dhanvijay, M. M., & Patil, S. C. Internet of Things: A survey of enabling technologies in healthcare and its applications. Computer Networks, 153, 113-131, 2019.

[14] Alam, M. M., Malik, H., Khan, M. I., Pardy, T., Kuusik, A., & Le Moullec, Y. A survey on the roles of communication technologies in IoT-based personalized healthcare applications. IEEE Access, 6, 36611-36631, 2018.

[15] Wang, H., Zhang, X., Xia, Y., & Wu, X. An intelligent blockchain-based access control framework with federated learning for genome-wide association studies. Computer Standards & Interfaces, 84, 103694, 2023.
[16] Shojafar, M., Mukherjee, M., Piuri, V., & Abawajy, J. Guest editorial: Security and privacy of federated learning solutions for industrial IoT applications. IEEE Transactions on Industrial Informatics, 18(5), 3519-3521, 2021.
[17] Mazzocca, C., Romandini, N., Colajanni, M., & Montanari, R. FRAMH: A Federated Learning Risk-Based Authorization Middleware for Healthcare. IEEE Transactions on Computational Social Systems, 2022.
[18] Bhansali, P. K., Hiran, D., Kothari, H., & Gulati, K. Cloud-based secure data storage and access control for the internet of medical things using federated learning. International Journal of Pervasive Computing and Communications, (ahead-of-print), 2022.
[19] Dhiman, G., Juneja, S., Mohafez, H., El-Bayoumy, I., Sharma, L. K., Hadizadeh, M., ... & Khandaker, M. U. Federated learning approach to protect healthcare data over big data scenario. Sustainability, 14(5), 2500, 2022.
[20] Jabal, A. A., Bertino, E., Lobo, J., Verma, D., Calo, S., & Russo, A. FLAP--A Federated Learning Framework for Attribute-based Access Control Policies. arXiv preprint arXiv:2010.09767, 2020.
[21] Ghimire, B., & Rawat, D. B. Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things. IEEE Internet of Things Journal, 2022.
[22] Savazzi, S., Nicoli, M., Bennis, M., Kianoush, S., & Barbieri, L. Opportunities of federated learning in connected, cooperative, and automated industrial systems. IEEE Communications Magazine, 59(2), 16-21, 2021.
[23] Geng, J., Kanwal, N., Jaatun, M. G., & Rong, C. DID-eFed: Facilitating Federated Learning as a Service with Decentralized Identities. In Evaluation and Assessment in Software Engineering (pp. 329-335), 2021.
[24] Alam, T., & Gupta, R. Federated Learning and Its Role in the Privacy Preservation of IoT Devices. Future Internet, 14(9), 246, 2022.
[25] Kim, T. Y., & Cho, S. B. Optimizing CNN-LSTM neural networks with PSO for anomalous query access control. Neurocomputing, 456, 666-677, 2021.
[26] Ye, X., Yu, Y., & Fu, L. Multi-Channel Opportunistic Access for Heterogeneous Networks Based on Deep Reinforcement Learning. IEEE Transactions on Wireless Communications, 21(2), 794-807, 2021.
[27] Otoum, Y., Liu, D., & Nayak, A. DL-IDS: a deep learning–based intrusion detection framework for securing IoT. Transactions on Emerging Telecommunications Technologies, 33(3), e3803, 2022.
[28] Kumar, A., Abhishek, K., Bhushan, B., & Chakraborty, C. Secure access control for manufacturing sector with application of ethereum blockchain. Peer-to-Peer Networking and Applications, 14(5), 3058-3074, 2021.
[29] Zhong, H., Zhou, Y., Zhang, Q., Xu, Y., & Cui, J. An efficient and outsourcing-supported attribute-based access control scheme for edge-enabled smart healthcare. Future Generation Computer Systems, 115, 486-496, 2021.
[30] Kumar, R., & Tripathi, R. Scalable and secure access control policy for healthcare system using blockchain and enhanced Bell–LaPadula model. Journal of Ambient Intelligence and Humanized Computing, 12(2), 2321-2338, 2021.
[31] Egala, B. S., Pradhan, A. K., Badarla, V., & Mohanty, S. P.Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control. IEEE Internet of Things Journal, 8(14), 11717-11731, 2021.
[32] Singh, A., & Chatterjee, K. LoBAC: A Secure Location-Based Access Control Model for E-Healthcare System. In Advances in Machine Learning and Computational Intelligence (pp. 621-628). Springer, Singapore, 2021.
[33] Younis, M., Lalouani, W., Lasla, N., Emokpae, L., & Abdallah, M. Blockchain-enabled and data-driven smart healthcare solution for secure and privacy-preserving data access. IEEE Systems Journal, 2021.
[34] Ghayvat, H., Pandya, S., Bhattacharya, P., Zuhair, M., Rashid, M., Hakak, S., & Dev, K. CP-BDHCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications. IEEE Journal of Biomedical and Health Informatics, 26(5), 1937-1948, 2021.
[35] Yaqoob, I., Salah, K., Jayaraman, R., & Al-Hammadi, Y. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. Neural Computing and Applications, 34(14), 11475-11490, 2022.
[36] Azad, M. A., Arshad, J., Mahmoud, S., Salah, K., & Imran, M.. A privacy-preserving framework for smart context-aware healthcare applications. Transactions on Emerging Telecommunications Technologies, 33(8), e3634, 2022.
[37] Balaji, N. V. An attack Resistant Privacy-Preserving Access Control Scheme for Outsourced E-pharma Data in Cloud. International Journal of Next-Generation Computing, 13(3), 2022.
[38] Tao, Q., & Cui, X. B-FLACS: blockchain-based flexible lightweight access control scheme for data sharing in cloud. Cluster Computing, 1-11, 2022.
[39] Pal, S., Dorri, A., & Jurdak, R. Blockchain for IoT access control: Recent trends and future research directions. Journal of Network and Computer Applications, 103371, 2022.
[40] Ghillani, D. Deep Learning and Artificial Intelligence Framework to Improve the Cyber Security. Authorea Preprints, 2022.
[41] Chinnasamy, P., & Deepalakshmi, P. HCAC-EHR: hybrid cryptographic access control for secure EHR retrieval in healthcare cloud. Journal of Ambient Intelligence and Humanized Computing, 13(2), 1001-1019, 2022.
[42] Astillo, P. V., Duguma, D. G., Park, H., Kim, J., Kim, B., & You, I. Federated intelligence of anomaly detection agent in IoTMD-enabled Diabetes Management Control System. Future Generation Computer Systems, 128, 395-405, 2022.
[43] Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., ... & He, B. A survey on federated learning systems: vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering, 2021.
[44] Karimi, L., Aldairi, M., Joshi, J., & Abdelhakim, M. An automatic attribute based access control policy extraction from access logs. IEEE Transactions on Dependable and Secure Computing, 2021.
[45] Hu, V. Machine Learning for Access Control Policy Verification (No. NIST Internal or Interagency Report (NISTIR) 8360 (Draft)). National Institute of Standards and Technology, 2021.

[46] Cotrini, C., Weghorn, T., & Basin, D. Mining ABAC rules from sparse logs. In 2018 IEEE European Symposium on Security and Privacy (EuroS&P) (pp. 31-46). IEEE, 2018.
[47] Amazon.com, "Amazon employee access challenge." Kaggle.
[48] Montanez, Ken, "Amazon access samples." UCI Machine Learning Repository: Amazon Access Samples Data Set.
[49] Rikhtechi, L., Rafe, V., & Rezakhani, A. Secured access control in security information and event management systems. Journal of Information Systems and Telecommunication, 9(33), 67-78, 2021.
[50] Rathna, R., Gladence, L. M., Cynthia, J. S., & Anu, V. M Energy efficient cross layer MAC protocol for wireless sensor networks in remote area monitoring applications. Journal of Information Systems and Telecommunication (JIST), 3(35), 207, 2021.

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 105-116
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

An Acoustic Echo Canceller using Moving Window to Track Energy Variations of Double-Talk-Detector

Mouldi Makdir[1*], Mohamed Bouamar[1,2], Mourad Benziane[2]

[1] Department of Electronics, Faculty of Technology, University of M'sila, M'sila, 28000, Algeria
[2] Laboratory of Analysis of Signals and Systems, M'sila, 28000, Algeria

Received: 21 Nov 2022/ Revised: 07 Oct 2023/ Accepted: 06 Dec 2023

Abstract

As a fundamental device in acoustic echo cancellation (AEC) systems, the echo canceller based on adaptive filters relies on the adaptive approximation of the echo-path. However, the adaptive filter must face the risk of divergence during the double-talk periods when the near-end is present. To solve this problem, the double-talk-detector (DTD) is often used to detect the double-talk periods and prevent the echo canceller from being disturbed by the other end of the speaker's signal. In this paper, we propose a DTD based on a new method that can detect quickly and track accurately double-talk periods. It is based on the sum of energies of the estimated echo and the microphone signals which is continuously compared to the error energy. A window that moves with time and tracks energy variations of the different input signals of the DTD represents a fundamental feature of the proposed method compared to several other methods based on correlation. The goal is to outperform conventional normalized cross-correlation (NCC) methods which are well-known in terms of small steady-state misalignment and stability of decision variable. In this work, the normalized least mean squares (NLMS) algorithm is used to update the filter coefficients along speech signals which are taken from the NOIZEUS database. Efficiency of the proposed method is particularly compared to the conventional Geigel algorithm and normalized cross-correlation method (NCC) that depends on the cross-correlation between the microphone signal and the error signal of AEC. Performance evaluation is confirmed by computer simulation.

Keywords: AEC; DTD; NLMS; NCC; Moving Window.

1- Introduction

The technique of acoustic echo cancellation (AEC) known for its interest in various applications of signal processing plays an important role in the field of telecommunications. The use of "hands-free" terminals allows maintaining the speaker's freedom of movement and ensuring the comfort of conversations. When the acoustic echo is present in a troublesome way, specific treatment must imperatively be implemented to preserve the quality of communication. The object of such a treatment is to estimate the acoustic echo between the received signal (signal sent in the loudspeaker) and the output of the room (signal picked up by the microphone) then to subtract an estimate of this output's signal without affecting the local speech signal in the case of double-talk (DT) [1,2]. This processing is done by using adaptive filtering where a double-talk-detector (DTD) is used to sense when the echo signal is corrupted by near-end signal. The role of this main function is to freeze adaptation of the filter coefficients when the near-end speech is present in order to avoid divergence of the adaptive algorithm [3-5].

Other methods based on combined adaptive filtering without DTD retain the advantages of both fast convergence rate and small steady-state misalignment but suffer from the same problems encountered in this field, such as abrupt changes in the acoustic echo-path, surrounding noise, and tracking capability. Indeed, they are complex and consume more computing time [6, 7].

In this work, we propose an efficient DTD to solve the problem provoked by the acoustic echo with the capability to improve speech intelligibility during telephone calls. To do this, a simulation will be started to allow a comparative study with two other methods [8-11].

The paper is organized as follows, in Section 2, the acoustic echo canceller with the proposed DTD is presented. In Sections 3 and 4, the used methods and the proposed one are formulated. The computational complexity is illustrated in Section 5. Simulation results are discussed in Section 6. Finally, the conclusion is given in Section 7.

✉ **Mouldi Makdir**
mouldi.makdir@univ-msila.dz

2- Acoustic Echo Cancellation

The acoustic echo canceller is used to remove the echo created due to the loudspeaker-microphone environment. We present in Fig.1 the structure of the device based on the new DT-detection method. In this case, the proposed DTD is controlled with three input signals where the energies of the estimated echo $\hat{y}(n)$ and the microphone signal $d(n)$ are continuously compared to the error energy of the signal $e(n)$.

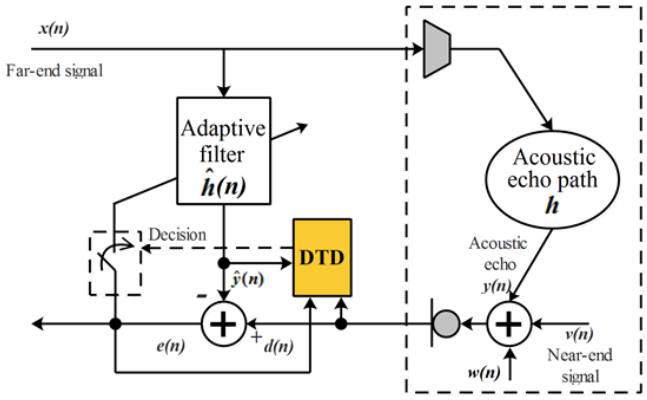


Fig. 1 Acoustic echo canceller with the proposed DTD

Note that the far-end vector signal $\mathbf{x}(n)$is filtered by the impulse response $\mathbf{h}$ modeling the room. At time n, the resulting signal (echo $y(n)$) is added to the near-end signal $v(n)$ and background noise $w(n)$ to give the corrupted microphone signal $d(n)$.
We have:

$$d(n) = y(n) + v(n) + w(n) \tag{1}$$

$\mathbf{x}(n)$is filtered through the impulse response $\mathbf{h}$ to get the echo signal:

$$y(n) = \mathbf{h}^T\mathbf{x}(n) \tag{2}$$

where:

$$\mathbf{x}(n) = [x(n) \quad x(n\text{-}1) \quad \dots \quad x(n\text{-}L+1)]^T$$
$$\mathbf{h} = [h_0 h_1 \quad \dots \quad h_{L-1}]^T$$

We have assumed that the length L of the vector signal $\mathbf{x}(n)$ is the same as the effective length of the echo-path $\mathbf{h}$.
At time n, the estimated echo $\hat{y}(n)$ is created by the convolution of the coefficients vector of adaptive filter $\hat{\mathbf{h}}(n)$ with the received input vector signal $\mathbf{x}(n)$.

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n\text{-}1)\,\mathbf{x}(n) \tag{3}$$

Where $\hat{\mathbf{h}}(n) = [\hat{h}_0(n) \quad \hat{h}_1(n) \quad \dots \quad \hat{h}_{L-1}(n)]^T$

The estimated echo signal is subtracted from the microphone signal, and the error signal is therefore given by:

$$e(n) = d(n) - \hat{y}(n) \tag{4}$$

Error signal which represents the error of the impulse response estimation is used in the adaptive algorithm to adapt the L coefficients of the filter $\hat{\mathbf{h}}$.
Several algorithms have been used to update the adaptive filter coefficients to converge to the optimal solution such as least mean squares (LMS), normalized least mean squares (NLMS), recursive least squares (RLS), and affine projection algorithms [4,12-15]. As an adaptive filtering algorithm that allows updating the filter coefficients, we use NLMS [16] to validate the proposed method. This is one, of the most used adaptive filtering algorithms which is defined by:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{\beta}{c + \mathbf{x}^T(n)\mathbf{x}(n)} e(n)\mathbf{x}(n) \tag{5}$$

Where $\hat{\mathbf{h}}(n+1)$ it is the next tap weight value, and $\hat{\mathbf{h}}(n)$ the current tap weight value of the adaptive filter. A constant β ($0 < \beta < 2$) controlling convergence is considered as a stabilization factor and a step size parameter used in updating the weight vector. The regularization parameter is a constant $c > 0$ that prevents division by zero [3, 12].
When the near-end signal is not present with any adaptive algorithm, the filter $\hat{\mathbf{h}}$ will quickly converge to an estimate of the echo-path $\mathbf{h}$ and this is the best way to cancel the echo. When the far-end signal is not present, or very small, the adaptation is stopped by the nature of the adaptive algorithm. When both signals are present, the near-end signal could disrupt the adaptation of filter $\hat{\mathbf{h}}$ and cause divergence. An effective DT-detection algorithm is used to stop adaptation of filter $\hat{\mathbf{h}}$ as fast as possible when the level of the near-end signal becomes appreciable in relation to the level of the far-end signal and to keep the adaptation going when the level of near-end signal is negligible. This is the case where it is important to use efficient DTD.

3- Double-Talk-Detection

DT-detection is used with an echo canceller to sense when echo signal is corrupted by near-end signal. Its role is to freeze the adaptation of the filter $\hat{\mathbf{h}}$ when near-end signal is present in order to avoid divergence of the adaptive algorithm. Typically, the DT-detection algorithm calculates a decision variable $\xi(n)$ and the DT is declared when $\xi(n)$ it is lower or upper than a threshold level T[10,11,17].
Methods based on DT-detection can be classified into two major categories, namely signal energy based and signal correlation based. Several methods such as cross-correlation (CC) [18-22], coherence, voice activity

detection, and fundamental frequency estimation have been proposed in the literature [23-26]. Methods based on cross-correlation between the far-end and error signals are then proposed. Moreover, approximate versions, such as a normalized cross-correlation (NCC), are developed but with a different combination of DTD input signals. Therefore, we will discuss in this study two of the most prominent methods in order to demonstrate their underlying ideas.

3-1- Geigel Method

A simple but elegant DT-detection algorithm was proposed by Geigel which is widely used for its easy implementation [8]. It is usually limited to network echo application where the echo level is typically 6 dB below that of far-end signal. It performs an amplitude level comparison between the maximum of a length L_G observation of $\mathbf{x}(n)$ and the microphone signal $d(n)$, where the decision variable is defined as :

$$\xi_G(n) = \frac{\max\left\{|x(n)|, |x(n-1)|, ..., |x(n-L_G+1)|\right\}}{|d(n)|} \quad (6)$$

L_G it is a constant that determines the number of past samples of the far-end signal that are used for the DT-detection. Decision is made by comparing $\xi_G(n)$ with a suitable threshold level T_G [19].

3-2- Cross-Correlation Method

The first method based on the cross-correlation between the far-end signal and the error signal is proposed by Hua Ye and Bo-Xiu Wu [9]. Some approximate versions as NCC are appeared in different articles where each method differs from the others in the DTD input signals [10,11,27]. Among these, we find one that depends on the cross-correlation between the microphone signal $d(n)$and the error signal $e(n)$ which we will use in this paper with the mentioned Geigel algorithm to compare them with the proposed method. Note that the performance of the proposed method in [11] is exactly similar to the best-known cross-correlation based DTD [10].
A statistical decision ξ_{NCC} of the NCC method is given by [11]:

$$\xi_{NCC}(n) = 1 - \frac{\hat{r}_{ed}(n)}{\hat{\sigma}_d^2(n)} \quad (7)$$

Where r_{ed} is the cross-correlation between $e(n)$ and $d(n)$, and σ_d^2 the variance of $d(n)$.

$\xi_{NCC(n)}$, it is based on estimates $\hat{r}_{ed}(n)$ and $\hat{\sigma}_d^{\,2}(n)$ which are found by using exponential weighting recursive estimation form [28, 29]:

$$\hat{r}_{ed}(n) = \lambda \hat{r}_{ed}(n-1) + (1-\lambda) e(n) d(n) \quad (8)$$

$$\hat{\sigma}_d^2(n) = \lambda \hat{\sigma}_d^2(n-1) + (1-\lambda) d(n) d(n) \quad (9)$$

Where λ is the exponential weighting factor ($0.9 < \lambda < 1$).
It should be noted that this method based on recursive estimation has a remarkable performance. However, it is significantly simpler and computationally very efficient. In addition, its main advantage is that only the maximum value of cross-correlation needs to be computed instead of computing the entire cross-correlation vector required by the other algorithms [11].
In this work, we propose to compare particularly the NCC method with the proposed one which is based on a moving temporal window used to track energy variations of three vector signals: error vector signal $\mathbf{e}(n)$, microphone vector signal $\mathbf{d}(n)$ and estimated vector signal $\hat{\mathbf{y}}(n)$.

4- Proposed Method

A fundamental feature of the proposed method compared to other ones is based on a window that moves with time to track energy variations of each input signal of the DTD. Three input signals to control the DTD are used where the sum of energies of the estimated echo and the microphone signals is continuously compared to the error signal energy.

We get the three input vector signals of the proposed DTD at time n as:

$$\mathbf{e}(n) = [e(n) \quad e(n\text{-}1) \quad \dots \quad e(n\text{-}N\text{+}1)]^T \quad (10)$$

$$\mathbf{d}(n) = [d(n) \quad d(n\text{-}1) \quad \dots \quad d(n\text{-}N\text{+}1)]^T \quad (11)$$

$$\hat{\mathbf{y}}(n) = [\hat{y}(n) \quad \hat{y}(n\text{-}1) \quad \dots \quad \hat{y}(n\text{-}N\text{+}1)]^T \quad (12)$$

Where N it is a constant length of the temporal window chosen to compute initial energy. It determines the number of past samples for each input vector signal of the DTD. From equation 4, we define the energy of the error vector signal as:

$$\|\mathbf{e}(n)\|^2 = \|\mathbf{d}(n) - \hat{\mathbf{y}}(n)\|^2 \quad (13)$$

$$\|\mathbf{e}(n)\|^2 = \|\mathbf{d}(n)\|^2 + \|\hat{\mathbf{y}}(n)\|^2 - 2\mathbf{d}^T(n)\hat{\mathbf{y}}(n) \quad (14)$$

With: $\|.\|$, the Euclidian norm of a vector.
Equation 15 can be defined as the decision variable ξ_{EE} (n) of the DTD.

$$\xi_{EE}(n) = \frac{\|\mathbf{e}(n)\|^2}{\|\mathbf{d}(n)\|^2 + \|\hat{\mathbf{y}}(n)\|^2} \quad (15)$$

With:

$0 < \xi_{EE}(n) < 1$ if $\{ \mathbf{d}^T(n)\ \hat{\mathbf{y}}(n)\ \} > 0$

$\xi_{EE}(n) > 1$ if $\{ \mathbf{d}^T(n)\ \hat{\mathbf{y}}(n)\ \} < 0$

- If $\mathbf{d}(n)$ and $\hat{\mathbf{y}}(n)$ are independents, $\xi_{EE}(n)$ =1, it is the case of orthogonality when DT is present.
- If $\mathbf{d}(n) = \hat{\mathbf{y}}(n)$, $\xi_{EE}(n) = 0$, it is the theoretical case of similarity when DT is not present.

Variations values of $\xi_{EE}(n) \geq 0$ reflect or not the presence of DT-situations. In Fig. 2, we show the variation range of the threshold levels (zones Z_0 and Z_1) where it will be judicious that a constant threshold level (T_{EE}), will be set initially in zone Z_0 to control the adaptive filter $\hat{\mathbf{h}}$.

The binary decision is then calculated as follows:

- if $\xi_{EE} > T_{EE}$, DT detected, the binary decision = 1, then no adaptation of the filter $\hat{\mathbf{h}}$;
- if $\xi_{EE} \leq T_{EE}$, DT not detected, the binary decision = 0, then adaptation of the filter $\hat{\mathbf{h}}$.

In practical cases or under hostile environments, the choice of a fixed threshold level with other methods will no longer be valid and must imperatively be replaced by an adaptive threshold [30]. However, the proposed method presents a nice property based on its ability to initially set one and only one fixed threshold level with $T_{EE} \approx 0$.When the far-end signal is present, the relation between the vector signals $\mathbf{d}(n)$ and $\hat{\mathbf{y}}(n)$ swings between two states : slightly correlated (when $\mathbf{v}(n) \neq 0$) and strongly correlated (when $\mathbf{v}(n) = 0$). It is considered that if the value of the threshold level T_{EE} is fixed in the zone Z_0, the better the correlation between the two vector signals ($\mathbf{d}(n)$ and $\hat{\mathbf{y}}(n)$) and the adaptation of the filter $\hat{\mathbf{h}}$ will be initiated.

Initial energies of the different input vector signals of the DTD are computed with a constant length N of the temporal window. The energy evolution of each vector signal is based on the preliminary calculation of an initial quantity of energy with a small number N of samples.

We have:

$$\|\mathbf{e}(j)\|^2 = \sum_{i=j}^{N+j-1} e^2(i) \quad (16)$$

$$\|\mathbf{d}(j)\|^2 = \sum_{i=j}^{N+j-1} d^2(i) \quad (17)$$

$$\|\hat{\mathbf{y}}(j)\|^2 = \sum_{i=j}^{N+j-1} \hat{y}^2(i) \quad (18)$$

Initially, the decision variable is calculated as:

$$\xi_{EE}(0) = \frac{\|\mathbf{e}(0)\|^2}{\|\mathbf{d}(0)\|^2 + \|\hat{\mathbf{y}}(0)\|^2} \quad (19)$$

When error signal $e(n)$ evolves with time, we get the following:

at time n=1

$$\|\mathbf{e}(1)\|^2 = \sum_{i=1}^{N} e^2(i) = \sum_{i=0}^{N-1} e^2(i) - e^2(0) + e^2(N)$$
$$= \|\mathbf{e}(0)\|^2 - e^2(0) + e^2(N) \quad (20)$$

at time n=2

$$\|\mathbf{e}(2)\|^2 = \sum_{i=2}^{N+1} e^2(i) = \sum_{i=1}^{N} e^2(i) - e^2(1) + e^2(N+1)$$
$$= \|\mathbf{e}(1)\|^2 - e^2(1) + e^2(N+1) \quad (21)$$
$$\vdots \qquad\qquad \vdots$$

at time n=k

$$\|\mathbf{e}(k)\|^2 = \sum_{i=k}^{N+k-1} e^2(i) = \sum_{i=k-1}^{N+k-2} e^2(i) - e^2(k-1) + e^2(N+k-1)$$
$$= \|\mathbf{e}(k-1)\|^2 - e^2(k-1) + e^2(N+k-1) \quad (22)$$

Idem, we get with $\mathbf{d}(k)$ and $\hat{\mathbf{y}}(k)$:

$$\|\mathbf{d}(k)\|^2 = \|\mathbf{d}(k-1)\|^2 - d^2(k-1) + d^2(N+k-1) \quad (23)$$

$$\|\hat{\mathbf{y}}(k)\|^2 = \|\hat{\mathbf{y}}(k-1)\|^2 - \hat{y}^2(k-1) + \hat{y}^2(N+k-1) \quad (24)$$

Decision variable that evolves continuously with time is given at k ($k > 0$) by:

$$\xi_{EE}(k) = \frac{\|\mathbf{e}(k-1)\|^2 - e^2(k-1) + e^2(N+k-1)}{\|\mathbf{d}(k-1)\|^2 - d^2(k-1) + d^2(N+k-1) + \|\hat{\mathbf{y}}(k-1)\|^2 - \hat{y}^2(k-1) + \hat{y}^2(N+k-1)} \quad (25)$$

5- Computational Complexity

As previously reported, energy evaluation of each input vector signal of the DTD is computed by using a temporal window initialized at the beginning with a constant length N. The calculation moves with time sample by sample and the decision variable is then evaluated on each time. In Fig. 3, we show an example of a moving temporal window which tracks energy variations of a signal.

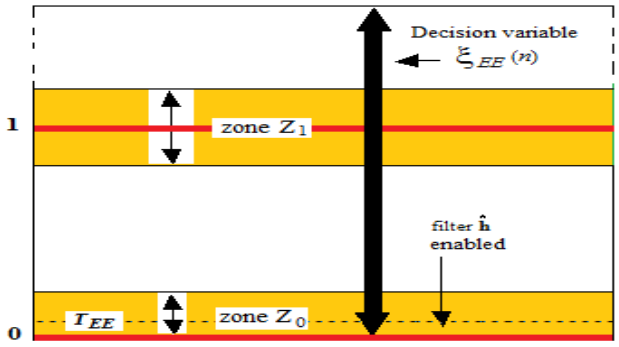


Fig. 2 Variation range of the threshold levels

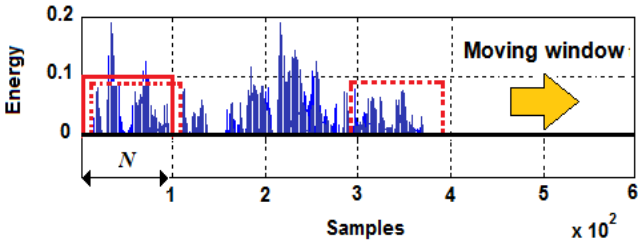


Fig. 3 Moving temporal window

It should be noted that the initial calculation of the energy performed on N samples for each input vector signal is done only at the beginning of the process. Thus, and as the evolution of these signals with time, the squared calculation will be done only on the new sample which replacing the oldest. Accordingly, by moving the window sample by sample, the total energy of each input signal will evolve continuously.

After N iterations, a first move of the window is thus achieved and the process works like a FIFO memory. The Fig. 4 shows an example of the first move of the initial window. At each time and for each input vector signal, we have one and only one squared sample computed. We require per iteration: 1 addition, 1 division and for each signal vector, 1 multiplication, 1 addition and 1 subtraction to compute the decision variable (i.e. 11 operations). A comparison between the previous and proposed method for the total number of computations per iteration is given in Table 1.

Table 1: Computational complexity per iteration

Method	Add	Sub	Mul	Div	Comp
Geigel	0	0	0	1	L_G-1
NCC	2	1	6	1	0
Proposed	4	3	3	1	0

The comparison indicates that Geigel method has a higher computational complexity. The algorithm depends directly on the tap-length L_G of the window used to calculate the maximum of **x**(n) samples. On the contrary, the proposed and NCC methods are independent regardless of this parameter. Furthermore, the proposed method remains faster than NCC with only three operations of multiplication per iteration. It appears that the proposed method can be considered more efficient for optimizing computation time.

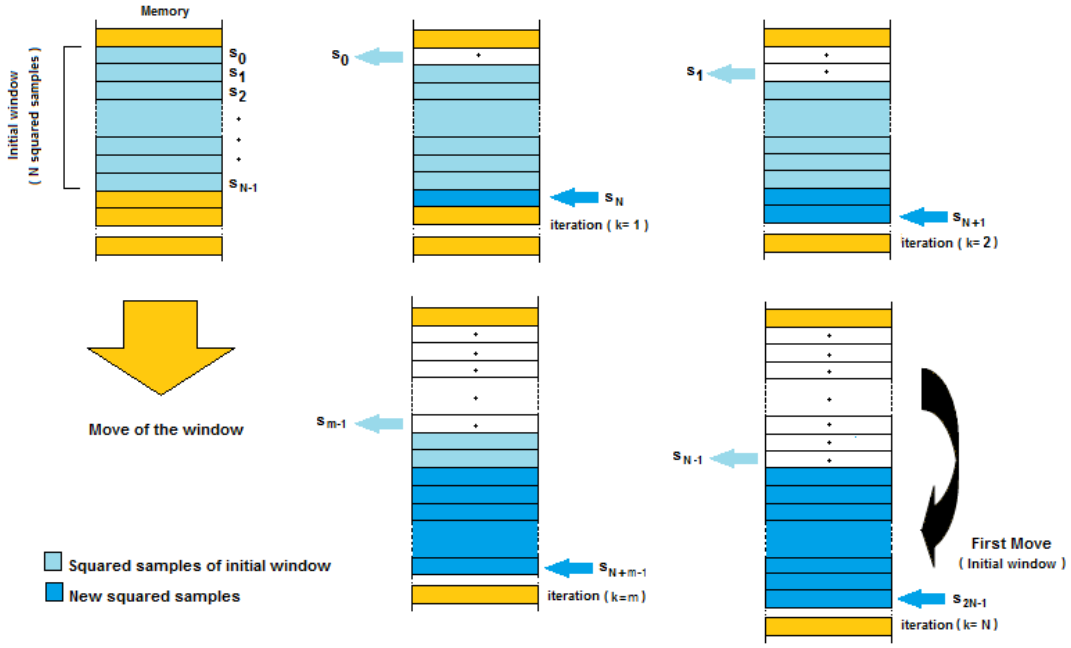


Fig. 4 First Move of the initial window based on FIFO technique

6- Simulation results

In this section, we evaluate the performances of the proposed method compared with Geigel and NCC using three different scenarios of speech signals (Sc1, Sc2, and Sc3) which are sampled at 8 kHz and issued from the NOIZEUS database. The echo model is based on the real impulse response with the length of echo-path L = 128 [31,32]. The three scenarios are presented in Fig. 5.

Three criteria for evaluating the performance of the proposed method are used: Misalignment, Echo Return Loss Enhancement (ERLE) and the probability of miss detection (P_m) [33,34].

The criteria are given as follows:

$$Misalignment(dB) = 10\log_{10}\left[\frac{\left\|\hat{\mathbf{h}}(n)-\mathbf{h}\right\|^2}{\left\|\mathbf{h}\right\|^2}\right] \quad (26)$$

$$ERLE(dB) = 10\log_{10}\left(\frac{E\left\{\left|d(n)^2\right|\right\}}{E\left\{\left|e(n)^2\right|\right\}}\right) \quad (27)$$

$$P_m = 1-\frac{\sum_{n=1}^{M}\bar{x}(n)\bar{v}(n)\phi(n)}{\sum_{n=1}^{M}\bar{x}(n)\bar{v}(n)} \quad (28)$$

where:

P_m is defined as the probability of detection failure when DT is present.

$\bar{x}(n)$ is the voice activity detection of far-end signal *x(n)*.

$\bar{v}(n)$ is the voice activity detection of near-end signal *v(n)*.

$\phi(n)$ is the binary decision of the DTD.

In the first step tests, a comparison of the different methods is performed with the background noise *w(n)=0*.Parameters used to update the adaptive filter $\hat{\mathbf{h}}$ are summarized in Table 2. Geigel, NCC and the proposed method have been performed respectively with the best parameter values selected for the scenario Sc1 and indicated in Table 3.

Table 2: Parameter values of AEC adaptive filtering

Parameter	Value
β	0.3
C	5.10^{-6}
L	128

Table 3: Parameter values selected for the different methods

Method	Parameter	Value
Geigel	T_G	0.8
	L_G	128
NCC	T_{NCC}	0.982
	λ	0.95
Proposed	T_{EE}	0.001
	N	40

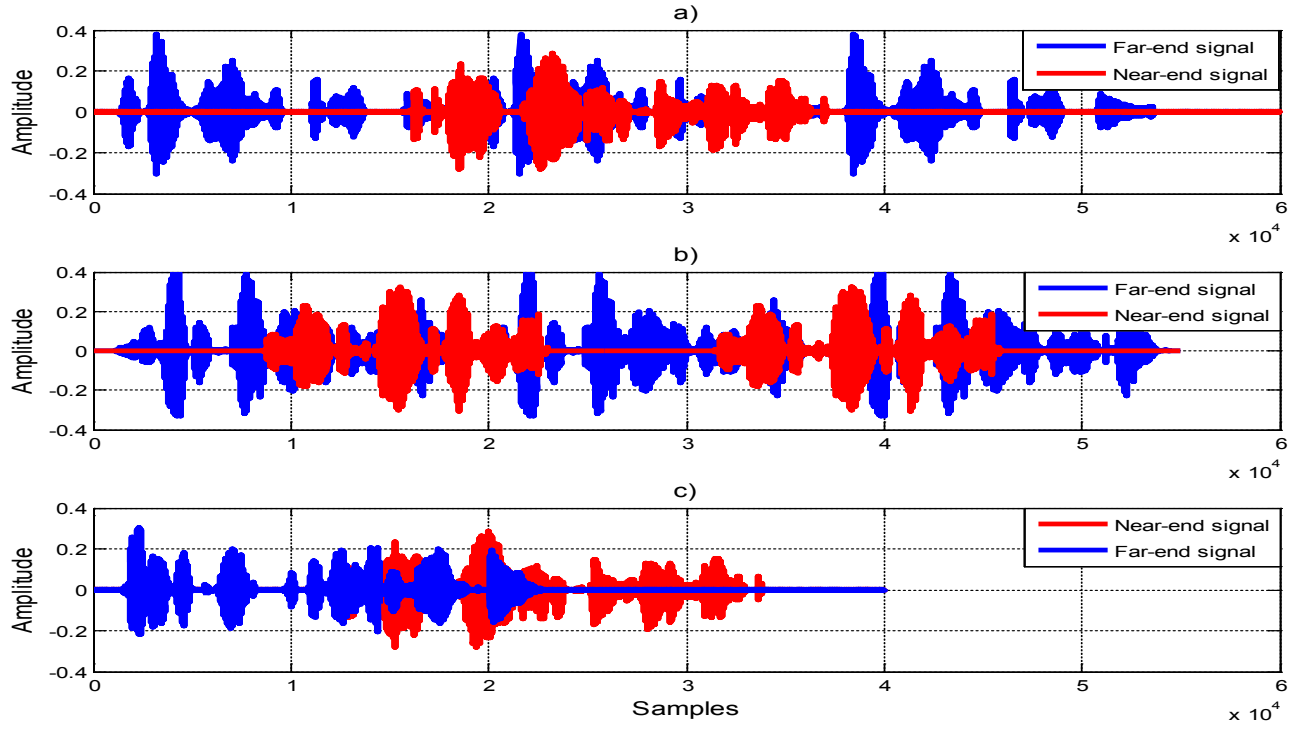


Fig. 5 Speech signals of the three Scenarios, a) Scenario Sc1, b) Scenario Sc2, c) Scenario Sc3.

Table 4: Parameter values of ERLE of the different methods with the three Scenarios.

Scenario	Geigel			NCC			Proposed		
	Peak	Average	Min	Peak	Average	Min	Peak	Average	Min
Sc1	46,49	17,14	-0,77	51,38	17,77	-0,33	52,19	19,96	-0,65
Sc2	60,50	19,69	-1,34	62,80	14,60	-0,47	67,96	23,97	-0,51
Sc3	47,98	11,26	-0,32	45,80	10,63	-0,96	45,85	11,47	-0,98

The ERLE criterion is considered to be one of the most used criteria in performance measurements of AEC algorithms. Recommendation G.131 of the International Telecommunications Union (ITU) requires an attenuation of more than 40 dB in the absence of double-talk [35]. The obtained results with the above scenarios presented in Table 4 and Fig. 6, confirm the superiority of the proposed method with peak values of an echo attenuation more than (52 dB for Sc1, 67 dB for Sc2, and 45 dB for Sc3).

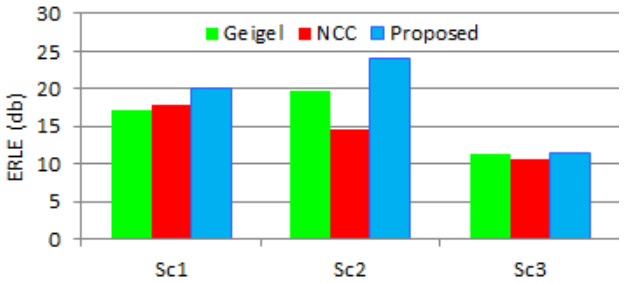


Fig. 6 Evolution of ERLE average of the different methods with the three scenarios

In Fig. 7, we compare the performance of the different methods in terms of misalignment. We remark that in the single-talk and before the apparition of the DT-period, the filter $\hat{\mathbf{h}}$ converges. Indeed, the proposed method maintained the constancy of the filter coefficients as soon as a DT-period occurred, whereas the NCC does false detection with a relative divergence. The Geigel method has detected too late the occurrence of DT-period with more divergence of the filter $\hat{\mathbf{h}}$. Therefore, the proposed method shows its superiority in terms of small steady-state misalignment and stability of decision variable.

In order to validate the proposed method concerning the choice of the parameter value of T_{EE} indicated in Table 3, we propose to illustrate in Fig. 8 the evolution of the decision variable $\xi_{EE}(n)$ obtained with the above scenarios.

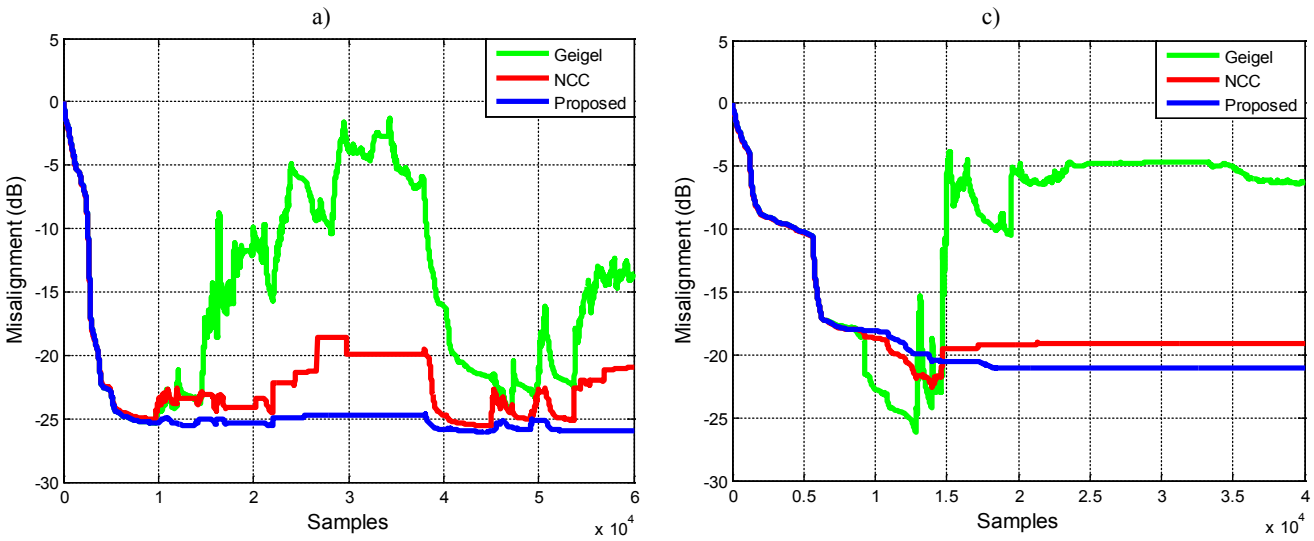


Fig. 7 Misalignment evaluation of the different methods with the three scenarios: a) Sc1, b) Sc2, c) Sc3.

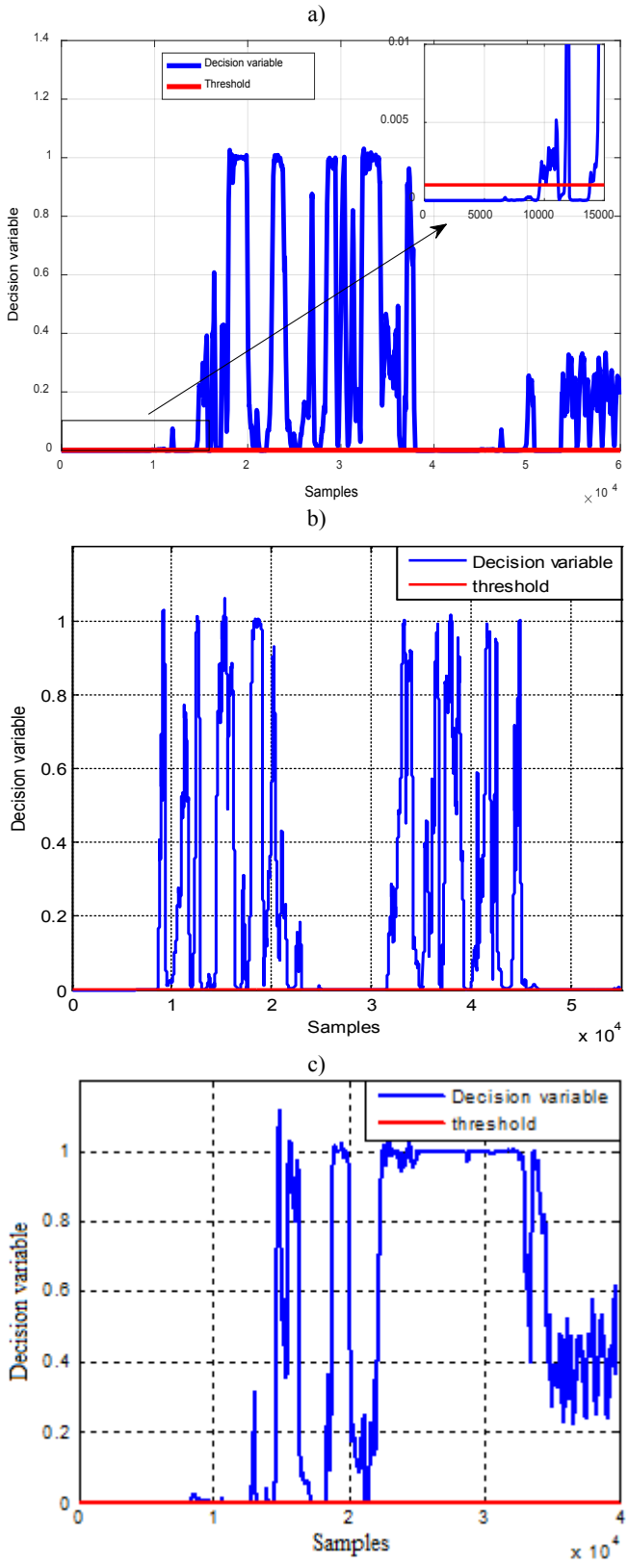


Fig. 8 Evolution of the decision variable of the proposed method with the three scenarios: a) Sc1, b) Sc2, c) Sc3.

It appears that with the scenarios (Sc1, Sc2 and Sc3) the decision variable $\xi_{EE}(n)$ displays a very close value to zero during single-talk periods and confirms the choice of a constant threshold level fixed in the zone Z_0.

To assess the impact of the fixed threshold level on the performance of the above methods, we show in Fig.9 the misalignments obtained with different threshold levels for the two scenarios (Sc2 and Sc3). It can be seen that the proposed method shows for an appropriate threshold level (T_{EE} = 0.001) initially set in zone Z_0 with scenario Sc1, leads to a result without degradation of the misalignment performance in scenarios Sc2 and Sc3. On the other hand, with Geigel and NCC methods, it can be seen that the threshold levels chosen with scenario Sc1 have been replaced by other more adequate threshold levels thus maintaining the performance of the corresponding misalignments. Therefore, we consider that for the proposed method, the T_{EE} threshold level initially set for a given scenario will also be valid with any other scenarios. Rather, Geigel and NCC methods will require an adaptive threshold level to maintain misalignment performance.

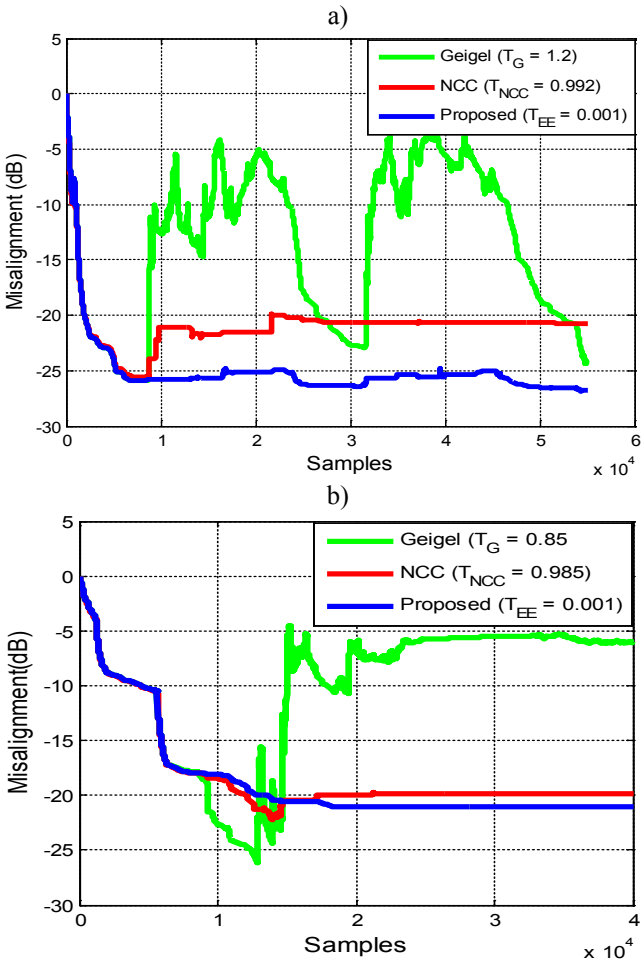


Fig. 9 Misalignment evaluation of the different methods with variable threshold, a) Sc2, b) Sc3.

In Fig. 10, we propose to evaluate with scenario Sc1 the impact of the length N on the misalignment of the proposed method. The results show misalignments with different values of N which demonstrate that a better performance is obtained with an appropriate set of N (N<100). We confirm that the preliminary calculation of energies requires a small number of samples and a reduced length N of the moving temporal window justifies a good tracking capability.

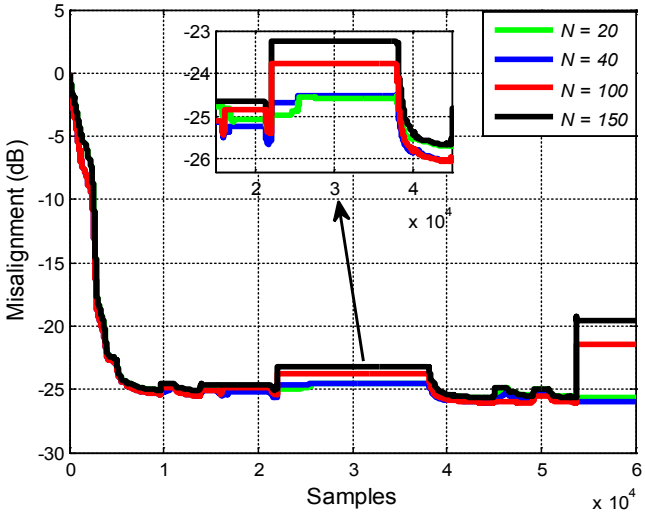


Fig. 10 Misalignment evaluation of the proposed method with different values of the length N

In order to evaluate ERLE and misalignment in a noisy environment ($w(n) \neq 0$),an independent white Gaussian noise is added to the echo signal of the scenario Sc1 with different signal-to-noise ratio (SNR) in the period between 6400 and 60000samples. Note that a constant noise with SNR = 50 dB is added only to the first 6400 input samples.

The SNR is defined as:

$$SNR(dB) = 10\log_{10}\left\{\frac{E\left[|y(n)|^2\right]}{E\left[|w(n)|^2\right]}\right\} \quad (29)$$

Near-end and Far-end signals are used with different levels of near-end-to-far-end ratio (NFR), which is calculated as:

$$NFR(dB) = 10\log_{10}\left\{\frac{E\left[|v(n)|^2\right]}{E\left[|x(n)|^2\right]}\right\} \quad (30)$$

We show in Table 5 parameter values of ERLE in a noisy environment obtained from the different methods with scenario Sc1. The results demonstrate a better and an appropriate ERLE values performed by the proposed method compared to Geigel and NCC.

Fig.11, illustrates clearly the superiority of ERLE average values obtained by the proposed method in a noisy environment.

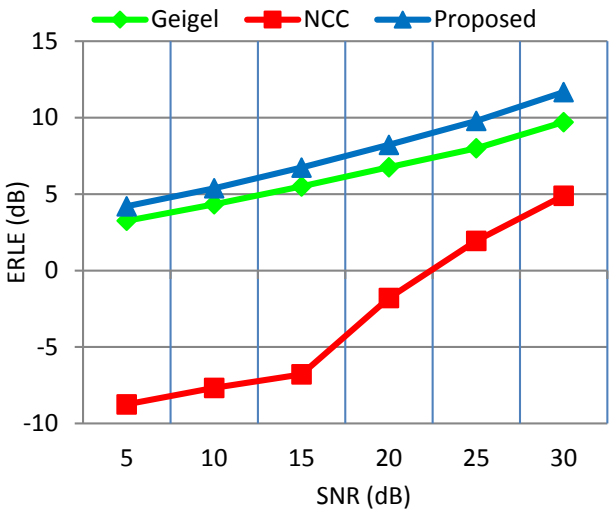


Fig. 11 Evolution of ERLE average for the different methods in a noisy environment

Table 5: Parameter values of ERLE of the different methods in a noisy environment.

SNR (dB)	Geigel			NCC			Proposed		
	Peak	Average	Min	Peak	Average	Min	Peak	Average	Min
5	38,46	3,26	-1,36	38,46	-8,75	-30,54	38,46	4,20	-0,90
10	38,46	4,34	-1,19	38,46	-7,67	-32,05	38,46	5,38	-0,72
15	38,46	5,50	-1,41	38,46	-6,79	-31,90	38,46	6,74	-0,55
20	38,46	6,76	-1,10	38,46	-1,79	-23,13	38,46	8,22	-0,50
25	38,46	8,00	-1,18	38,46	1,95	-9,94	38,46	9,79	-0,57
30	38,46	9,71	-1,16	38,46	4,89	-7,92	40,80	11,67	-0,59

Misalignment evaluation in a noisy environment with scenario Sc1 is illustrated in Fig. 12. The results show that the proposed method presents good performances in terms of misalignment and minimizing false detection in the DT-situation. Robustness against additive noise of the proposed method is clearly appeared compared to other ones.

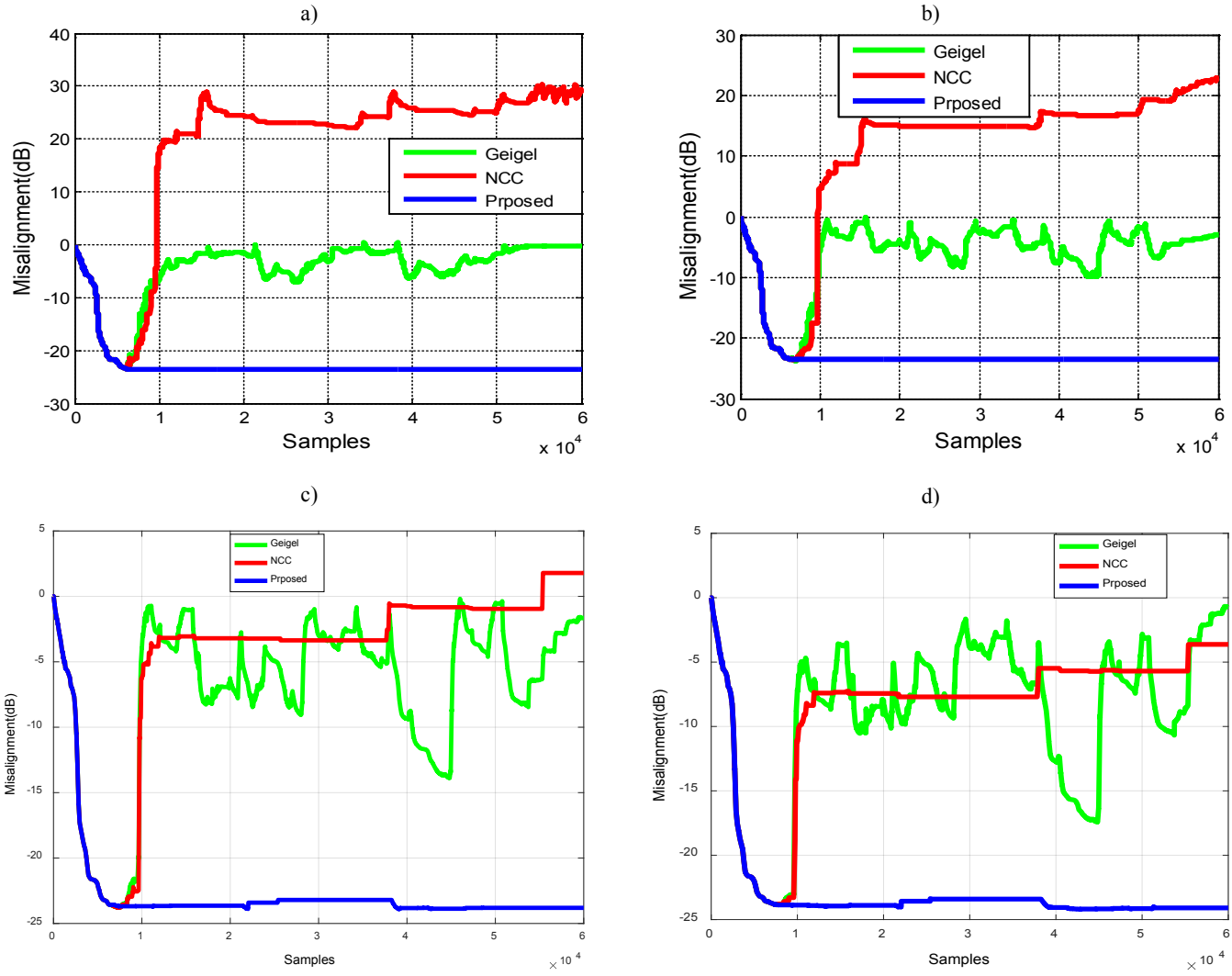


Fig. 12 Misalignment evaluation of the different methods in a noisy environment, a) SNR=5 dB, b) SNR=15 dB, c) SNR= 25 dB, d) SNR= 30 dB

To simulate the change in the echo-path, we increase the gain of the acoustic channel by 10 at sample 31000. The obtained results with scenario Sc1 are shown in Fig. 13. They demonstrate a good tracking capability by the proposed method which can distinguish between the near-end signal and an abrupt change of the acoustic channel.

Objective performance evaluation based on the probability of missed detection P_m is presented in Fig. 14. It is calculated with SNR = 20 dB as a function of NFR values varying between -10 dB and 20 dB. The used threshold for each method is chosen to give a probability of false detection $P_f = 0.2$ which is defined as the probability of declaring detection when DT does not exist. It is calculated without the near-end signal as:

$$P_f = \frac{1}{M} \sum_{n=1}^{M} \bar{x}(n)\phi(n) \tag{31}$$

The obtained result demonstrates that the proposed method is better than Geigel and NCC in terms of the probability of missed detection when NFR varies more than -10 dB.

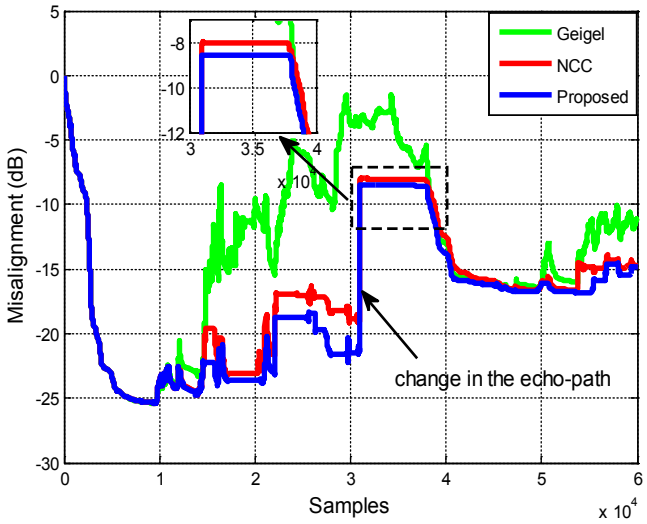


Fig. 13 Misalignment evaluation with a change in the echo-path

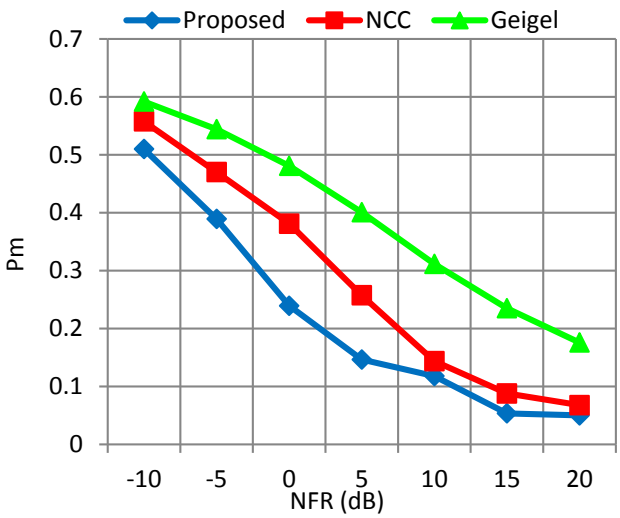


Fig. 14 Probability of missed detection

7- Conclusion

In this paper, we have presented a new and efficient method used for AEC systems where the main purpose is to halt quickly and accurately the update filter coefficients during the DT-periods. The method is based on a moving temporal window that tracks variations of the error energy compared to the sum of energies of the estimated echo and the microphone signals. We consider that the decision variable based on a window that moves with time to track variations of the error energy improves the distinguishing capability between far-end and near-end speech signals. Computer simulation has demonstrated the superiority of the proposed method in terms of small steady-state misalignment, high ERLE, and robustness against the additive white noise and abrupt change in the echo-path. It has also presented improvement in terms of minimizing the number of miss detection and false alarm with no variable threshold level. As an algorithm performed with FIFO technique, the proposed method can be considered also efficient for optimizing computation time. It is significantly simpler and has the capability to outperform conventional NCC methods. Further work remains necessary to compare it with other recent methods.

References

[1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, S. L. Gay, "Advances in network and acoustic echo cancellation. Digital Signal Processing," Springer, Berlin, Heidelberg, 2001.

[2] M. M. Sondhi, "An adaptive echo canceler," The Bell Syst, Technical journal, Vol. 46, No. 3, 1967, pp. 497–511.

[3] J. M. Gil-Cacho, "Adaptive filtering algorithms for Acoustic Echo Cancellation and Acoustic feedback control in speech communication applications," PhD. Thesis, University of Belgium Ku Leuven, 2013.

[4] S. Haykin, "Adaptive filter Theory," Prentice-Hall, Inc, Upper Saddle River, NJ, USA, 1996.

[5] J. Benesty, T. Gänsler, "Audio signal processing for next generation multimedia communication systems," Kluwer Academic Publishers, 2004.

[6] F. Huang, J. Zhang, S. Zhang, "Combined-step size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments," IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 63, No. 5, 2015, pp. 493-497.

[7] Y. R. Chien, J. Li-You, "Convex combined adaptive filtering algorithm for acoustic echo cancellation in hostile environments," IEEE Access, Vol. 6, 2018, pp. 16138-16148.

[8] D. Duttweiler, "A twelve-channel digital echo canceler," IEEE Transactions on Communications, Vol. 26, No. 5, 1978, pp. 647-653.

[9] H. Ye, B. X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," IEEE Transactions on Communications, Vol. 39, No. 39, 1991, pp. 1542-1545.

[10] J. Benesty, D. R. Morgan, J. H. Cho, "A new class of double-talk detectors based on cross-correlation," IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 2, 2000, pp. 168-172.

[11] M. A. Iqbal, J. W. Stokes, S. L. Grant, "Normalized double-talk detection based on microphone and AEC error cross- correlation," IEEE International Conference on Multimedia and Expo, 2007, pp. 360-363.

[12] P. S. R. Diniz, "Adaptive Filtering Algorithms and Practical Implementation," Springer, 2013.

[13] M. Hajiabadi, "Acoustic Noise Cancellation Using an Adaptive Algorithm Based on Correntropy Criterion and Zero Norm Regularization," JIST Journal of Information

Systems and Telecommunication, Vol. 3, No. 3, 2015, pp. 150-156.

[14] Hun, Choi.Hyeon-Deok, Bae, "Subband Affine Projection Algorithm for Acoustic Echo Cancellation System," EURASIP Journal on Advances in Signal Processing, 2007, pp. 1-12.

[15] B. H. Yang, "An adaptive filtering algorithm for non-Gaussian signals in alpha-stable distribution," Traitement du Signal, Vol. 37, No. 1, 2020, pp.69-75.

[16] S. Hannah, D. Samiappan , R. Kumar , A. Anand, A. Kar, "Variable tap-length non-parametric variable step-size NLMS adaptive filtering algorithm for acoustic echo cancellation," Applied Acoustics, Vol. 159, 2020.

[17] M. Hamidia, A. Amrouche, "A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation," Digital Signal Processing, Vol. 60, 2017, pp. 99-112.

[18] V. Thien-An, H. Ding, M. Bouchard, "A survey of double-talk detection schemes for echo cancellation applications," Canadian Acoustics, Vol. 32, No. 3, 2004, pp. 144-145.

[19] M. Benziane, M. Bouamar, M. Makdir, "Doubletalk detection based on enhanced Geigel algorithm for acoustic echo cancellation," In 2018 6th International Conference on Control Engineering & Information Technology (CEIT), 2018, pp. 1-5.

[20] T. Gänsler, J. Benesty, "A frequency-domain double-talk detector based on a normalized cross-correlation vector," Signal Processing, Vol. 81, No 8, 2001, pp. 1783–1787.

[21] J. Benesty, T. Gänsler, "A multichannel acoustic echo canceler double-talk detector based on a normalized cross-correlation matrix*," European Transactions on Telecommunications, Vol. 13, No 2, 2002, pp. 95–101.

[22] T. Gänsler, J. Benesty, "The fast normalized cross-correlation double-talk detector," Signal Process, Vol. 86, No. 6, 2006, pp. 1124–1139.

[23] T. Gansler, M. Hansson, C.J.Ivarsson, G. Salomonsson, "A double-talk detector based on coherence,"IEEE Transactions on Communications, Vol. 44, No. 11, 1996, pp. 1421-1427.

[24] H. Bao, Y. Yang, J. Liu, X. Ba, Q. Yuan, "A robust algorithm of double talk detection based on voice activity detection," Proc. Inter. conf. on Audio Language and Image Processing, 2010, pp. 12–15.

[25] S. Cecchi, L. Romoli, F. Piazza, "Multichannel Double-Talk Detector based on Fundamental Frequency Estimation," IEEE Signal Processing Letters, Vol. 23, No. 1, 2016, pp. 94-97.

[26] Y. Zhenhai, F. Yang, J. Yang, "Optimum step-size control for a variable step-size stereo acoustic echo canceller in the frequency domain," Speech Communication, Vol. 124, 2020, pp. 21–27.

[27] S. J. Park, C. G. Cho, C. Lee, D. H. Youn, S. H. Park, "Integrated echo and noise canceller for hands free applications," IEEE Transactions on circuits and systems, Part II, Analog and Digital Signal Processing, Vol. 49, No. 3, 2002, pp. 188-195.

[28] Y. Hua, "Adaptive filter theory and applications," PhD. Thesis, South-East university, China, 1989

[29] Honig, M.L., Messerschmitt, D.G., "Adaptive Filters," Kluwer, 1984.

[30] M. Benziane, M. Bouamar, M. Makdir, "Simple and Efficient Double-Talk-Detector for Acoustic Echo Cancellation," Traitement du signal, Vol. 37, No. 4, 2020, pp. 585-592.

[31] ITU-T. "Digital Network Echo Cancellers," Recommendation G.168, International Telecommunication Union; Geneva, 2007.

[32] Y. Hu, P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," Speech Communication , Vol. 49, No. 7, 2007, pp. 588-601.

[33] H. Wonchul, K. Taehwan, B. Keunsung, "Robust double-talk detection in the acoustic echo canceller using normalized error signal power," Proc. ISSPA'07.UAE, 2007, pp. 1-4

[34] J.H. Cho, D.R. Morgan, J. Benesty., "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 6, 1999, pp. 718–724.

[35] ITU-T. "Digital Network Echo Cancellers," Recommendation G.131, International Telecommunication Union; Geneva, 2003.

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 117-126
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

An Aspect-Level Sentiment Analysis Based on LDA Topic Modeling

Sina Dami[1*], Ramin Alimardani[1]

[1].Department of Computer Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran

Received: 15 Jun 2022/ Revised: 04 May 2024/ Accepted: 08 Jun 2024

Abstract

Sentiment analysis is a process through which the beliefs, sentiments, allusions, behaviors, and tendencies in a written language are analyzed using Natural Language Processing (NLP) techniques. This process essentially comprises of discovering and understanding people's positive or negative sentiments regarding a product or entity in the text. The increased significance of sentiments analysis has coincided with the growth in social media such as surveys, blogs, Twitter, etc. The present study takes advantage of the topic modeling approach based on latent Dirichlet allocation (LDA) to extract and represent the thematic features as well as a support vector machine (SVM) to classify and analyze sentiments at the aspect level. LDA seeks to extract latent topics by observing all the texts, which is accomplished by assigning the probability of each word being attributed to each topic. The important features that represent the thematic aspect of the text are extracted and fed to a support vector machine for classification through this approach. SVM is an extremely powerful classification algorithm that provides the possibility to separate complex data from one another accurately by mapping the data to a space with much larger aspects and creating an optimal hyperplane. Empirical data on real datasets indicate that the proposed model is promising and performs better compared to the baseline methods in terms of precision (with 89.78% on average), recall (with 78.92% on average), and F-measure (with 83.50% on average).

Keywords: Natural Language Processing; Sentiment Analysis; Aspect-Level; Topic Modeling; LDA.

1- Introduction

Sentiment analysis or opinion mining is a research field aimed at expressing the behavior, sentiments, opinions, and analysis of various individuals regarding entities and their features. These entities can be goods, services, organizations, other individuals, events, and topics that have to do with information recovery and knowledge extraction from the test through data mining and natural language processing [1]. Text information in the world is divided into two groups of facts and sentiments [2]. Facts are real phrases about the entities, events, and their features, whereas sentiments are mental phrases that indicate the sentimental opinions of people and their thoughts and beliefs regarding an entity, event, or one of their features. So far, ample research has been conducted on factual information. For instance, information extraction [3], textual implication [4], text summarization [5], classification [6], clustering [7], and many other applications can be mentioned in natural language processing and text mining sciences [8]. In contrast, few studies have been conducted on sentimental information. One of the most important reasons for the shortage in studies on texts containing sentiments and beliefs compared to texts containing facts is the existence of much less sentimental information, particularly before the expansion of the worldwide web. Aside from the facts, beliefs and sentiments are quite significant too since we strive to know others' opinions whenever we want to decide on action [9].

We would ask the opinions of friends and families when deciding on the expansion of the worldwide web, and organizations and firms used to need public surveys, questionnaires, or interviews when they needed the opinions of the public regarding their goods or services [10]. The number of texts and pages containing sentiments started accelerating with the emergence of the worldwide web. In other words, the internet has changed how people's sentiments, opinions, and perspectives change so that people can reflect their opinions on commercial pages, internet groups, and blogs [11].

These online opinions of people make for a vast resource of assessable information that can be used for many applications. Basically, opinions and sentiments can be

✉ **Sina Dami**
Dami@wtiau.ac.ir

analyzed at three levels of granularity, including document, sentence and word. The sentiment expressed in an entire text, for example, review sentence or document, is called overall sentiment. The task of analyzing overall sentiments of texts is normally formulated as classification problem, e.g., classifying a review sentence or document into positive or negative sentiment. Then, different types of machine learning approaches trained using different levels of granularity (features) have been applied for overall sentiment analysis. The existing methods at each of these three levels can also be categorized into three groups including supervised learning, semi-supervised learning, and unsupervised learning.

The present study has adopted a hybrid approach combining supervised and unsupervised learning to perform topic modeling of texts and classify sentiments, respectively. Latent Dirichlet Allocation (LDA) was used for this purpose to model several latent variables (titles) in a set of texts encompassing words. A Support Vector Machine (SVM) was also used to classify and analyze sentiments in both positive and negative aspects. SVM learning precision and data classification in social media platforms can be enhanced using LDA for semantic extraction of the topics at the level or words' roots. On the basis of this hypothesis, following two research questions were identified: 1) What is the overall performance of aspect-level sentiment analysis based on LDA topic modeling? 2) How efficient is hybrid machine learning approach to extract aspects for sentiment analysis?

The main contributions of this research are as follows:

- Develop a topic modeling approach for aspect-based sentiment analysis applicable to any product or service.
- Specifically, identify the topics of books, electronics, video games, cell phones, luxury beauty and group their attributes into aspects.
- Adopt a hybrid approach combining supervised and unsupervised learning to perform aspect-based sentiment analysis.

2- 2- Literature Review

Researchers have mainly studied the process of sentiment analysis in three grained levels so far including the document (text) level, the sentence level, and the aspect level [12]. A commented text document (.g. a critic on a product) is classified as a text indicating a completely positive or a completely negative opinion. This type of classification considers the while text as one unit of information and assumes that the desired text is a commented text containing opinions regarding a specific entity (e.g. a certain phone). Sentiment classification at the sentence level [13] classifies the single sentences in a text; however, one cannot assume that a comment has been made in each sentence. The conventional method is to first divide the sentences into commenting and non-commenting sentences; a process called subjective classification. Then, the commenting sentences are classified into the sentences expressing positive comments and the ones expressing negative comments. Sentence-level sentiment classification can also be formulated as a three-class categorization so that each sentence can be classified as positive, negative, or neutral.

Compared to the document-level and sentence-level analysis, aspect-level analysis or sentiment analysis based on the aspect [14] is considerably more fine-grained. This analytical process extracts and summarizes users' opinions regarding the aspects/features of entities, which are called goals as well. For instance, the goal of aspect-based sentiment analysis in the case of a product's critics is to summarize the positive and negative opinions regarding the various aspects of the product, whether the overall opinion regarding the product is positive or negative. The main task of aspect-based analysis includes several sub-tasks including aspect extraction, entity extraction, and classification of aspect sentiments. For instance, in the case of the sentence "IPhone's audio quality is excellent but its battery is no good", the task of aspect extraction is to identify "iPhone" as the entity, and the aspect extraction must recognize that "audio quality" and "battery" are two distinct aspects. Aspect-level sentiment classification must also classify the sentiment expressed regarding Iphone's audio quality as a positive sentiment and the one regarding Iphone's battery as a negative sentiment. It must be noted that aspect and entity extraction are combined in many algorithms to make it work easier, and are called sentiment/opinion goal extraction or aspect extraction overall.

In user review mining, the approaches based on topic modeling and Latent Dirichlet Allocation (LDA) are important techniques used to extract the aspect of a product in aspect-based sentiment analysis [15]. The LDA approach has been proposed to address the problems and issues of LSA and PLSA algorithms [16]. LDA has been used in many sentiment analysis research works. In [17], LDA was used to understand the public response to COVID19 in Weibo. The authors collected 719,570 posts from the Weibo website using a web crawler and analyzed the data using text extraction techniques such as LDA topic modeling and sentiment analysis. Some of the results of this study indicated that in response to the COVID19, people learned about it, expressed their support for frontline workers and active individuals, give each other spiritual support, and expressed their concerns regarding life and economic revival when it some to preventive measures. Moreover, sentiment analysis indicated that the country's media and social media influencers help each other in posting positive sentiment information.

In [18], a new method is proposed to investigate the electronic reputation and negative sentiments regarding a tourism destination (Morocco in TripAdvisor) based on LDA. This study investigated around 39,216 TripAdvisor reviews from various attractions and places in Morocco to extract the latent aspects and dimensions in the reviews of tourists that have visited Morocco using LDA. Moreover, many studies using an adaptation of LDA for short texts have also been published, in which case the existing methods must be developed considering the problem of data scatter and the lack of synchronous patterns in short texts. In [15], an LDA-based method for aspect-level sentiment analysis of user reviews with short texts is proposed. The proposed method for aspect-level sentiment analysis was called the Sentence Segment LDA (SS-LDA). SS-LDA is a new adaptation of the LDA algorithm for product aspect extraction. Empirical results of examinations on some datasets revealed that SS-LDA is highly competitive in product aspect extraction. A similar work [19] was also performed social networks and micro-blogs, notably during the COVID-19 pandemic. In this work, an aspect-oriented sentiment classification was proposed using a combination of the prior knowledge topic model algorithm (SA-LDA), automatic labelling (SentiWordNet) and ensemble method (Stacking). Experimental results have shown that the proposed SA-LDA outperformed the standard LDA.

Venugopalan and Gupta [20] proposed an unsupervised approach for aspect term extraction, a guided Latent Dirichlet Allocation (LDA) model that uses minimal aspect seed words from each aspect category to guide the model in identifying the hidden topics of interest to the user. The guided LDA model is enhanced by guiding inputs using regular expressions based on linguistic rules. The model is further enhanced by multiple pruning strategies, including a BERT based semantic filter, which incorporates semantics to strengthen situations where co-occurrence statistics might fail to serve as a differentiator. The work has been evaluated on the restaurant domain of SemEval 2014, 2015 and 2016 datasets and has reported an acceptable evaluation.

Chen et al. [21] also combined user information and product information for classification but carried this out through sentence-level and word-level consideration that can account for both user priorities and product features at both sentence and word levels. Similarly, Dou [22] used a deep memory network to collect user and product information. The proposed model can be divided into two separate sections. Long Short-Term Memory (LSTM) is used in the first section to learn the display of a text. A deep memory network made up of several computational layers is used in the second section to predict the critical ranking for each text.

Mahadevaswamy and Swathi [23] investigated a technical review of sentiment analysis using a Bidirectional Long Short-Term Memory (LSTM) network. This network is deep and capable of leveraging long-term dependencies by bringing memory in the network for performing better analysis. Edara et al. [24] presented a deep learning model with LSTM network as an alternative to the classical sentiment analysis models.

Iparraguirre-Villanueva et al. [25] proposed an architecture to find out what people think about Monkeypox disease. They used a hybrid model based on CNN and LSTM architecture to determine the classification accuracy. Mohbey et al. [26] also proposed a hybrid architecture model based on CNN-LSTM models to find out people's feelings regarding Monkeypox epidemic. Their research goal was to investigate how the common sense about the Monkeypox disease to help politician in understanding of how the common views the epidemic, more deeply.

Although Recursive Neural Network with Long Short-Term Memory (LSTM) has been among the most successful techniques in many fields [27-30], but some research [13] and [14] indicate that the Support Vector Machine (SVM) approach performs better than RNN deep learning approach in terms of aspect-level sentiment analysis. Aurangzeb et al. [31] proposed an ensemble method based on the Support Vector Machine (SVM) for aspect-level sentiment analysis. Their method comprised of taking advantage of an evolutionary approach based on the Genetic Algorithm (GA) combined with the power of the SVM algorithm to examine multi-label text data. Empirical results on seven datasets (medical, hotel, movies, automobiles, proteins, birds, emotions, and news) demonstrated that using the SVM-GA algorithm outperformed many state-of-the-art algorithms such as Bayesian probability models [32] MLP and CNN neural networks [33], and multi-component learning methods [33, 35]. Therefore, this study also recommended an approach based on support vector machines combined with the advantages of LDA-based topic modeling techniques.

3- Proposed Method

Figure 1 demonstrates the process of the proposed method. The input data entering the system undergo preprocessing in the preprocessing stage, and the Eigenvalues are normalized. Then, topic modeling followed by reduction of the input data dimensions is carried out using the LDA algorithm to make the calculation more accurate and reduce calculation time. LDA observes all the words in the text, assigns the probability of each word belonging to each topic, and creates topics made out of words close to one another. Through this process, the redundant features that are not required in the analysis are eliminated, and only the features among the input data that affect the analysis remain. Then, training data extracted from the previous stage is used to build a support vector machine to

train the system so that it can learn the textual sentiment analysis pattern. Thus, the build model will in fact be the base for this sentiment analysis. At the next stage, the learning model created in the previous stage is used to perform the sentiment analysis based on input data and determine what the result of textual sentiment analysis is. Eventually, we assess the extent to which textual sentiment analysis has been conducted accurately given the outputs of the sentiment analysis system, and obtain the respective assessment metrics. The details of the proposed process are discussed in the following.

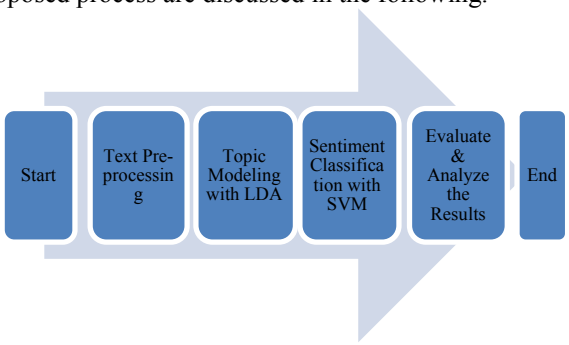


Fig. 1 The process of the proposed method

Algorithm 1 explains the details of how to use the proposed method.

Algorithm 1. Aspect-Level Sentiment Analysis Based on LDA Topic Modeling

```
Input: Words w ∈ documents d
Output: Sentiment
pre-processed text ← Pre-processing (w);
Topics ← LDA (pre-processed text);
Aspects ← Domain aspects;
for each Topic do
    Topic-Aspect mapping;
end
for each Aspect do
    for each Sentence words w in d do
        if w contains Topic words then
            Add Sentence words to Aspect sentences;
        else
            Skip Sentence words w;
        end
    end
end
Sentiment Model ← SVMTrain (Aspect sentences);
Sentiment Classification Score ← SVMTest (Sentiment Model);
for each Aspect Sentences in d do
    if Sentiment Classification Score > 0 then
        Sentiment ← Positive;
    else
        if Sentiment Classification Score < 0 then
            Sentiment ← Negative;
        else
            Sentiment ← Neutral;
        end
    end
end
```

3-1- Text Preprocessing

Contrary to structured data, textual data are not easily accessible, so we have to use a process to extract the features out of textual data. One way to do so is to consider each word as a feature and find a criterion for the presence or absence of the word in a sentence or the document. This technique is called the Bag-of-Word (BoW). The first step to creating the BoW is to convert each document into a feature vector so that each vector demonstrates the words in each document. Term Frequency-Inverse Document Frequency (TF-IDF) is the conventional method to determine the importance of the words in this mode. Before any analysis, the TF-IDF data must be normalized. This section discusses how data are normalized. Suppose we have the set X with specific values as mentioned in the following equation:

$$X = \{ X_1 . X_2 . \dots . X_n \}$$

Maximum and minimum members of the set are defined as Eqs. (1) and (2).

$$Min(X) = r \mid r \in X \wedge \forall s \in X : r \leq s \quad (1)$$

$$Max(X) = r \mid r \in X \wedge \forall s \in X : r \geq s \quad (2)$$

The set of the normalized values of each member of X normalized to fall between the values of a and b are calculated as Eq. (3) shows.

$$Norm(X) = \{ a + (b - a) * (X_i - Min(X) / (Max(X) - Min(X)) \mid X_i \in X \wedge 1 \leq i \leq n \} \quad (3)$$

Thus, the data in the set will become more suitable for the respective analysis and comparisons.

3-2- Topic Modeling of Text using LDA

Latent Dirichlet Allocation (LDA) is an unsupervised technique for the extraction of thematic information from a set of documents without labeled data. The main idea behind LDA is that documents are presented as a random combination of latent topics, each topic being the probability distribution of the words. Fig. 2 illustrates a graphic LDA model. In this figure, the nodes are random variables, the edges are the conditional relationships between the variables, and the rectangles are the iteration of the sampling steps throughout the production process by the number shown in the lower right corner of the

rectangle. For instance, the inner rectangle which contains the random variables of z and w is repeated N_d times of D various documents. The variables hatched in the figure are observed variables and the non-hatched ones are the latent variables of the model.

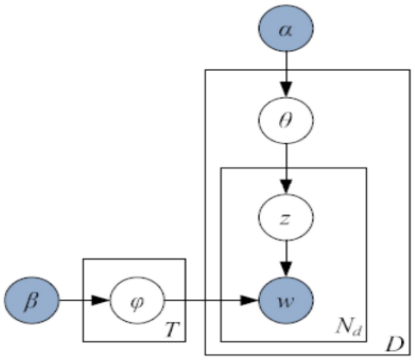


Fig. 2 The LDA structure in the form of a graphic probability model

LDA assumes that textual documents display various topics, i.e. they are made up of words that belong to various topics, and the ratio of the topics in the document varies. We can classify the document into a specific topic considering these ratios. For this purpose, a fixed set of words are considered as the glossary. The LDA method assumes that each topic is a distribution on this set of words; i.e. the words that are from one topic have a high probability in that topic. We assume that these topics are already specified. Now, we produce the words for each document among the available documents through the following two steps:

1. We randomly select the probability distribution on the topics;
2. For each word in the document:
   2.1. We specify a topic randomly using the probability distribution from the previous stage;
   2.2. We select a word from the glossary randomly given the specified probability distribution.

This probability model reflects how many topics each document is made up of. The first stage of this process demonstrates that various topics have different shares in one text. The second part of the second stage also indicates that each word in each document has been extracted from one of the topics while the first part of the second stage emphasizes that the topic has been selected from the probability distribution of topics on the documents. It must be mentioned that all documents include the same set of topics in this method, but each document incorporates different ratios of the topics.

Fig. 3 demonstrates the LDA model. M represents the number of texts and N represents the number of words in each text. The parameters of the model include:

α The Dirichlet prior distribution for the titles for each text

β Dirichlet prior distribution for the distribution of words for each title

θ_i The distribution of the titles for the ith text.

φ_k The distribution of words for the kth title

z_{ij} The latent variables of the jth word in the ith text

w_{ij} The jth word in the ith text

V the number of words

φ The $K \times V$ matrix of words' distribution for each title

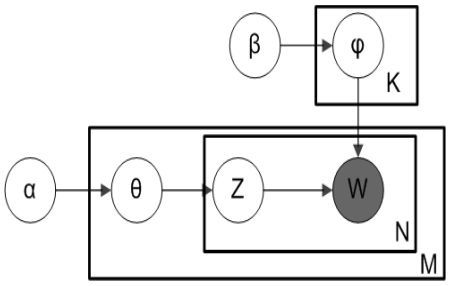


Fig. 3 The LDA model display

Only w_{ij} variables are observed and the rest are latent variables. Now, total data can be created based on the latent variables as follows:

Selecting the $\theta_i \sim Dir(\alpha)$ Dirichlet distribution for each $i \in \{1, \dots, M\}$

Selecting the $\varphi_k \sim Dir(\beta)$ Dirichlet distribution for each $k \in \{1, \dots, K\}$

For each w_{ij}:

Selecting the title of $z_{ij} \sim Multinominal(\theta_i)$

Selecting the words $w_{ij} \sim Multinominal(\varphi_{z_{ij}})$

More formally, Eq. (4) is used to calculate the word distribution according to the document.

$$p(w_i|d) = \sum_{j=1}^{K} p(w_i|z_j)p(z_j|d) \quad (4)$$

3-3- Sentiment Classification using SVM

Support vector machine is among the relatively new methods that have indicated good performance over the recent years compared to the older classification methods. The basis of the SVM classifier is linear data classification in which we try to select the line with a higher confidence margin. Solving the equation to find the optimal line for the data is performed through quadratic programming (QP) methods that are known methods used for solving constrained problems. The φ function is used to take the data to a space with a much higher dimension before the linear division so that the machine and classify highly complex data. To solve problems with extremely high dimensions using these methods, the Lagrangian duality theorem is used to minimize the problem into its dual form where a simpler function called the Kernel function which is the vector multiplication of the φ function instead of the

complex *Phi* function that takes us to a space with high dimensionality.

Suppose x is the input vector in a space with m dimensions and has been transferred to the news feature space of M using the base function $\varphi_j(x), j = 1,\ldots,M$. Thus, each next m input vector of xi, $i = 1,\ldots,n$ (n is the number of samples) will be converted into a new feature vector $\varphi(x_i) = [\varphi_1(x_i), \varphi_2(x_i), \ldots, \varphi_M(x_i)]^T$. Then, the support vector separator is designed based on what was mentioned in the previous sections. The (nonlinear) function in the new feature space is created as $\hat{f}(x) = \varphi(x)^T \hat{w} + \hat{w}_0$ (the equation can be simplified assuming $\forall x : \varphi_0(x) = 1$ as the bias which is multiplied by $\hat{w}_0$). The separator is $\hat{G}(x) = sign\ \hat{f}(x)$ as it used to be. The dimensions of the new feature space can be considered extremely large or even infinite, and calculations and losing generalizability are what things that cause restrictions in this field if the base functions are not considered adequately. For instance, consider Fig. 4). This dataset cannot be simply classified with a line in the input space. Mapping to a space with greater dimensions is used in such cases.

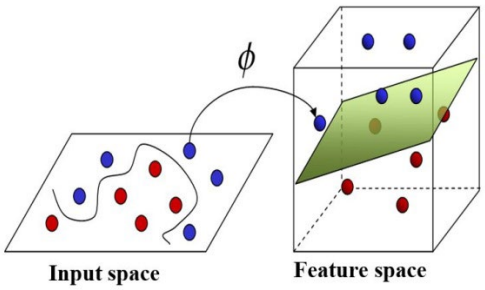


Fig. 4 Mapping a dataset to a space with greater dimensions

Thus, SVM tries to receive the input space with small dimensions and turn it into a space with greater dimensions using a method called the Kernel trick. This conversion turns an inseparable problem into a separable problem. These functions are called the Kernel functions. Kernel functions are pretty useful in nonlinear separation problems such as textual sentiment classification.

Although Kernel methods implicitly work in large spaces, it can be demonstrated that increasing the number of dimensions dies not to reduce their accuracy since the statistical learning theory has confirmed that the Kernel method's generalizability is ultimately dependent on the number of the samples classified incorrectly at the training stage. Thus, the selection of the Kernel function is the most significant issue in SVM. Many methods and principles such as Diffusion kernel, Fisher kernel, String kernel, etc. have been introduced for this purpose, and research is being carried out to obtain the Kernel matrix from the available data. In practice, a lower-degree polynomial Kernel or a Radial Base Function (RBF) kernel with an acceptable with is a good starting point. An SVM with RBF Kernel which has been used in the present study (Eq. (5)) which is quite close to RBF neural networks with Radial Base centers that are automatically selected for SVM.

$$K(X_i, X_j) = e^{-\|X_i - X_j\|^2 / 2\sigma^2} \quad (5)$$

4- Experimental Result

To evaluate the proposed method, the Amazon Review Dataset was used, which released in 2018 [36]. This Dataset includes reviews (ratings, summaries, text, time, helpfulness votes), product metadata (descriptions, category information, price, brand, and image features), and links. The data are available at https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/.

A real sample of customer review is shown in Fig. 5. As observable, the customer review consists of five important aspects:

- Rating: User rating of the product on a scale of 1 to 5.
- Summary: The title of the review
- Review text: The actual content of the review.
- Review time: The real time of the review (raw).
- Helpfulness: The number of people who found the review useful.

These aspects will help us comprehend and analyze the reviews to classify sentiments.

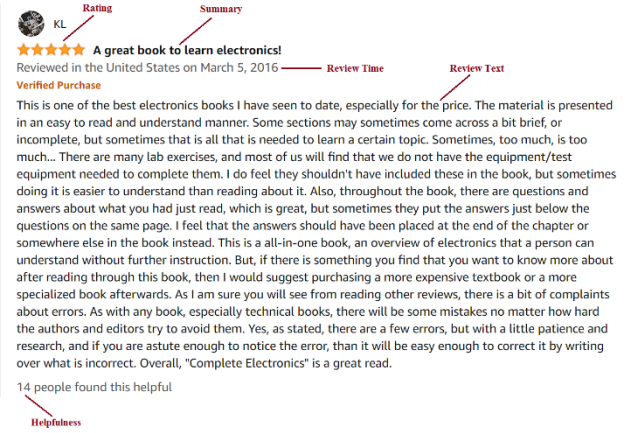


Fig. 5 Real Amazon customer review sample

The structure of the data is in JSON format as follows:

{
"reviewerID": "A2SUAM1J3GNN3B",
"asin": "0000013714",
"reviewerName": "J. McDonald",
"helpful": [2, 3],

"reviewText": "I bought this for my husband who plays the piano. He is having a wonderful time playing these old hymns. The music is at times hard to read because we think the book was published for singing from more than playing from. Great purchase though!",
"overall": 5.0,
"summary": "Heavenly Highway Hymns",
"unixReviewTime": 1252800000,
"reviewTime": "09 13, 2009"
}

We selected 5 per-category datasets, which includes 5-core reviews and product metadata for each category, mentioned in Table 1.

Table 1: Dataset description

Per-category	***Description***
Books	This category is about the reviews of books from Amazon[1].
Electronics	Electronics is a review dataset [37] collected from the Electronics category on Amazon with Clothing as an auxiliary category.
Cell Phones and Accessories	This category is about Amazon reviews predictions of Cell Phones and Accessories.
Luxury Beauty	This category performs sentiment analysis on Amazon reviews for Luxury Beauty products.
Video Games	This category is about classification and topic analysis of video game reviews, trained on Amazon user reviews.

[1] https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews

RapidMiner software was used to implement the proposed model. The parameters of SVM were set as Table 2.

Table 2: Parameter setting

Parameter	***Value***	***Description***
C	10	It is the regularization parameter, C, of the error term
kernel	rbf	It specifies the kernel type to be used in the algorithm
degree	3	It is the degree of the polynomial kernel function ('poly') and is ignored by all other kernels
gamma	auto	It is the kernel coefficient for 'rbf', 'poly', 'sigmoid'. If gamma is 'auto', then 1/n_features will be used instead

10-fold cross-validation was used to evaluate the five groups in the dataset. This method splits each dataset into 10 random sections and considers nine sections as the training and the remaining one section as the testing set each time. Then, it implements the proposed algorithms 10 times, calculates the evaluation criteria, obtains their mean, and reports it as the final output.

The two common evaluation criteria in sentiment analysis include precision and recall, which are quite applicable and tangible in the evaluation of various data mining algorithms. Precision and recall are defined in Eqs. (6) and (7), respectively:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Where TP (True Positive) represents the number of samples that have been correctly assigned to the positive class, FP (False Positive) indicates the number of samples that have been incorrectly assigned to the positive class, and FN (False Negative) represents the number of samples that have been incorrectly assigned to the negative class. It must be mentioned that positive and negative classes are the two positive and negative modes considered for the present data in the problem of sentiment classification. It can be demonstrated that every multiclass problem can easily be converted into a two-class problem by considering one class as positive and the others as negative each time. Thus, the calculations can be performed easily for the three classes of positive, negative, and neutral as well.

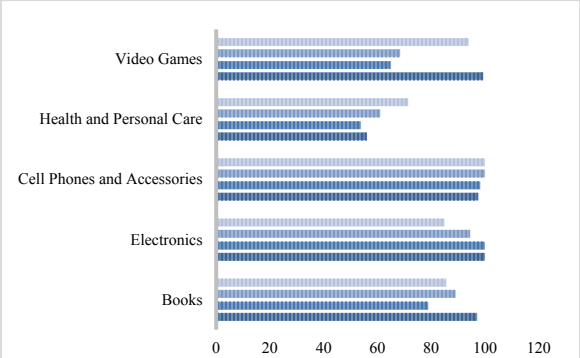
In addition to the two mentioned criteria, there are other criteria called the F-measure which is calculated based on the harmonic mean of precision and recall as indicated in Eq. (8).

$$F_\beta = \frac{(1+\beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{8}$$

One specific case of this parameter is the F-Score which equals F1 for β = 1 (Eq. (9)).

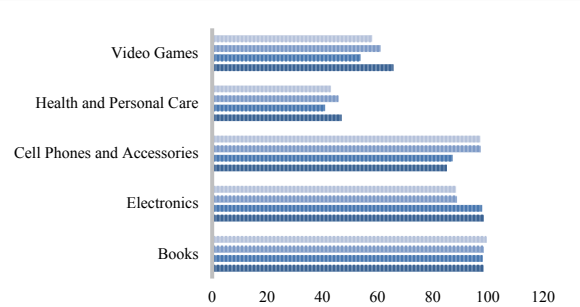
$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

These criteria can be used to present the results of the evaluation. To conduct a comparative analysis on the proposed method, its performance was compared to the three algorithms of Bagging [38], RNN-GRU [39], and LSTM-CRF [40]. Figures 6-8 demonstrate the comparison of precision, recall, and F-measure divided by the groups using all mentioned methods. As can be observed, the percentage of the mentioned criteria generally indicated better performance for the proposed method in all datasets which reveals that the proposed method is promising.



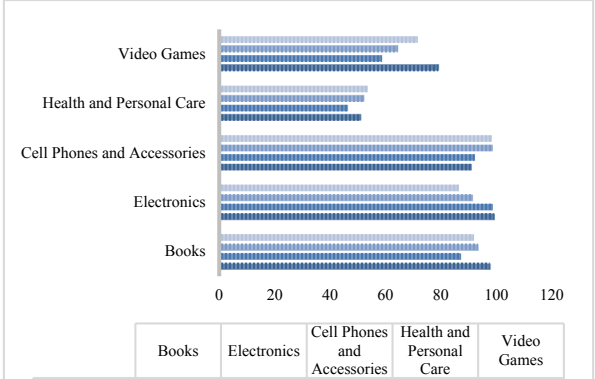
	Books	Electronics	Cell Phones and Accessories	Health and Personal Care	Video Games
Bagging	85.4	84.7	99.7	71.3	93.7
RNN-GRU	88.9	94.3	99.8	60.9	68.3
LSTM-CRF	78.6	99.6	97.9	53.6	64.7
LDA-SVM	96.8	99.7	97.3	56	99.1

Fig. 6 Comparison between the proposed method and other methods' precision (%)



	Books	Electronics	Cell Phones and Accessories	Health and Personal Care	Video Games
Bagging	99.4	88.2	97	42.9	57.9
RNN-GRU	98.2	88.5	97.1	45.7	60.9
LSTM-CRF	97.9	97.7	87	40.9	53.7
LDA-SVM	98.4	98.5	85.1	46.9	65.7

Fig. 7 Comparison between the proposed method and other methods' recall (%)



	Books	Electronics	Cell Phones and Accessories	Health and Personal Care	Video Games
Bagging	91.87	86.41	98.33	53.57	71.57
RNN-GRU	93.32	91.31	98.43	52.22	64.39
LSTM-CRF	87.19	98.64	92.13	46.4	58.69
LDA-SVM	97.59	99.1	90.79	51.05	79.01

Fig. 8 Comparison between the proposed method and other methods' F-measure (%)

Apart from the advantages of support vector machine's robustness, the use of topic modeling using the LDA algorithm improves precision, recall, and F-measure compared to the other methods. The proposed method had a favorable performance over the other methods for all cases of data with various observation probabilities, which clearly reveals its excellent performance. Although the proposed method has an insignificantly lower performance compared to the baseline method in a few cases, it leads to the best result and is the best method in most cases.
Looking at Fig. 6-8, we found that using LDA topic modeling improves the performance of aspect-level sentiment. Returning to the question raised at the beginning of this study, it is now possible to state that using the hybrid machine learning approach is efficient to extract aspects for sentiment analysis.

5- Conclusions and Recommendations

The present study proposed a new ethod for textual sentiment analysis using the approach of topic modeling based on Latent Dirichlet Allocation (LDA) combined with a support vector machine. In the process of topic modeling, each text comprises various topics and each topic includes various words. The LDA algorithm observes all these texts and tries to create topics made up of words that are semantically close to one another by assigning the probability of belonging to each topic to each word. Aside from developing the topics, it makes connections between them and the texts in the dataset. The important features that represent the thematic aspect of the text are extracted through this method and are fed to a support vector machine. The support vector machine is an extremely powerful classification algorithm that provides

the possibility to accurately separate complex data from one another by mapping them to a space with extremely higher dimensions. As the results of the evaluation indicated, this approach was revealed to have higher precision, recall, and F-measure compared to the rival methods.

To extend the proposed method, deep learning techniques combined with the advantages of topic modeling can be used for deep extraction and display of the features and model long-term dependencies inherent in the text. Moreover, other semantic approaches such as semantic role labeling and the use of Ontology can also replace or be combined with topic modeling.

References

[1] Luo F, Li C, Cao Z. Affective-feature-based sentiment analysis using SVM classifier. In2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD) 2016 May 4 (pp. 276-281). IEEE.

[2] Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. InThirty-second AAAI conference on artificial intelligence 2018 Apr 26.

[3] Dami S. News Events Prediction Based on Casual Inference in First-Order Logic (FOL). Journal of Soft Computing and Information Technology. 2016 Dec 21;5(4):11-25.

[4] Dami S, Barforoush AA, Shirazi H. News events prediction using Markov logic networks. Journal of Information Science. 2018 Feb;44(1):91-109.

[5] Rezaei A, Dami S, Daneshjoo P. Multi-document extractive text summarization via deep learning approach. In2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI) 2019 (pp. 680-685). IEEE.

[6] Dami S, Yahaghizadeh M. Efficient event prediction in an IOT environment based on LDA model and support vector machine. In2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) 2018 Feb (pp. 135-138). IEEE.

[7] Emami H, Dami S, Shirazi H. K-Harmonic Means Data Clustering With Imperialist Competitive Algorithm. University Politehnica of Bucharest-Scientific Bulletin, Series C: Electrical Engineering and Computer Science. 2015 Feb;77(7).

[8] García-Pablos A, Cuadros M, Rigau G. W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Systems with Applications. 2018 Jan 1;91:127-37.

[9] Goswami S, Nandi S, Chatterjee S. Sentiment analysis based potential customer base identification in social media. InContemporary Advances in Innovative and Applicable Information Technology 2019 (pp. 237-243). Springer, Singapore.

[10] Araque O, Corcuera-Platas I, Sánchez-Rada JF, Iglesias CA. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications. 2017 Jul 1;77:236-46.

[11] Zhou Q, Xu Z, Yen NY. User sentiment analysis based on social network information and its application in consumer reconstruction intention. Computers in Human Behavior. 2019 Nov 1;100:177-83.

[12] Moraes R, Valiati JF, Neto WP. Document-level sentiment classification: An empirical comparison between SVM and ANN. Expert Systems with Applications. 2013 Feb 1;40(2):621-33.

[13] Shirsat VS, Jagdale RS, Deshmukh SN. Sentence level sentiment identification and calculation from news articles using machine learning techniques. InComputing, Communication and Signal Processing 2019 (pp. 371-376). Springer, Singapore.

[14] Al-Smadi M, Qawasmeh O, Al-Ayyoub M, Jararweh Y, Gupta B. Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. Journal of computational science. 2018 Jul 1;27:386-93.

[15] Ozyurt B, Akcayol MA. A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. Expert Systems with Applications. 2021 Apr 15;168:114231.

[16] Parveen N, Santhi MV, Burra LR, Pellakuri V, Pellakuri H. Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK. Materials Today: Proceedings. 2021 Jan 6.

[17] Xie R, Chu SK, Chiu DK, Wang Y. Exploring public response to COVID-19 on Weibo with LDA topic modeling and sentiment analysis. Data and Information Management. 2021;5(1):86-99.

[18] Ali T, Marc B, Omar B, Soulaimane K, Larbi S. Exploring destination's negative e-reputation using aspect based sentiment analysis approach: Case of Marrakech destination on TripAdvisor. Tourism Management Perspectives. 2021 Oct 1;40:100892.

[19] AlGhamdi N, Khatoon S, Alshamari M. Multi-aspect oriented sentiment classification: Prior knowledge topic modelling and ensemble learning classifier approach. Applied Sciences. 2022 Apr 18;12(8):4066.

[20] Venugopalan M, Gupta D. An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. Knowledge-based systems. 2022 Jun 21;246:108668.

[21] Chen P, Sun Z, Bing L, Yang W. Recurrent attention network on memory for aspect sentiment analysis. InProceedings of the 2017 conference on empirical methods in natural language processing 2017 Sep (pp. 452-461).

[22] Dou ZY. Capturing user and product information for document level sentiment analysis with deep memory network. InProceedings of the 2017 Conference on Empirical Methods in Natural Language Processing 2017 Sep (pp. 521-526).

[23] Mahadevaswamy UB, Swathi P. Sentiment analysis using bidirectional LSTM network. Procedia Computer Science. 2023 Jan 1;218:45-56.

[24] Edara DC, Vanukuri LP, Sistla V, Kolli VK. Sentiment analysis and text categorization of cancer medical records with LSTM. Journal of Ambient Intelligence and Humanized Computing. 2023 May;14(5):5309-25.

[25] Iparraguirre-Villanueva O, Alvarez-Risco A, Herrera Salazar JL, Beltozar-Clemente S, Zapata-Paulini J, Yáñez JA, Cabanillas-Carbonell M. The public health contribution of sentiment analysis of Monkeypox tweets to detect polarities

using the CNN-LSTM model. Vaccines. 2023 Jan 31;11(2):312.
[26] Mohbey KK, Meena G, Kumar S, Lokesh K. A CNN-LSTM-Based Hybrid Deep Learning Approach for Sentiment Analysis on Monkeypox Tweets. New Generation Computing. 2023 Aug 14:1-9.
[27] Amin J, Sharif M, Raza M, Saba T, Sial R, Shad SA. Brain tumor detection: A long short-term memory (LSTM)-based learning model. Neural Computing and Applications. 2020 Oct;32(20):15965-73.
[28] Dami S, Yahaghizadeh M. Predicting cardiovascular events with deep learning approach in the context of the internet of things. Neural Computing and Applications. 2021 Jan 3:1-8.
[29] Dami S, Esterabi M. Predicting stock returns of Tehran exchange using LSTM neural network and feature engineering technique. Multimedia Tools and Applications. 2021 May;80(13):19947-70.
[30] Dami S. Internet of things-based health monitoring system for early detection of cardiovascular events during COVID-19 pandemic. World Journal of Clinical Cases. 2022 Sep 9;10(26):9207.
[31] Aurangzeb K, Ayub N, Alhussein M. Aspect Based Multi-Labeling Using SVM Based Ensembler. IEEE Access. 2021 Feb 1;9:26026-40.
[32] Yang Y, Jiang J. Adaptive bi-weighting toward automatic initialization and model selection for HMM-based hybrid meta-clustering ensembles. IEEE transactions on cybernetics. 2018 Mar 27;49(5):1657-68.
[33] Liao S, Wang J, Yu R, Sato K, Cheng Z. CNN for situations understanding based on sentiment analysis of twitter data. Procedia computer science. 2017 Jan 1;111:376-81.
[34] Kumar V, Pujari AK, Padmanabhan V, Kagita VR. Group preserving label embedding for multi-label classification. Pattern Recognition. 2019 Jun 1;90:23-34.
[35] Wu G, Zheng R, Tian Y, Liu D. Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification. Neural Networks. 2020 Feb 1;122:24-39.
[36] Ni J, Li J, McAuley J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. InProceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) 2019 Nov (pp. 188-197).
[37] Wan M, Ni J, Misra R, McAuley J. Addressing marketing bias in product recommendations. InProceedings of the 13th international conference on web search and data mining 2020 Jan 20 (pp. 618-626).
[38] Jahanbakhsh Gudakahriz S, Eftekhari Moghaddam AM, Mahmoudi F. An Experimental Study on Performance of Text Representation Models for Sentiment Analysis. Journal of Information Systems and Telecommunication (JIST). 2020 Jul;1(29):45.
[39] Chandra N, Ahuja L, Khatri SK, Monga H. Utilizing Gated Recurrent Units to Retain Long Term Dependencies with Recurrent Neural Network in Text Classification. Journal of Information Systems and Telecommunication (JIST). 2021 May;2(34):89.
[40] Xiong H, Yan H, Zeng Z, Wang B. Dependency Parsing and Bidirectional LSTM-CRF for Aspect-level Sentiment Analysis of Chinese. InJIST (Workshops & Posters) 2018 (pp. 90-93).

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 127-137
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

A Comparison Analysis of Conventional Classifiers and Deep Learning Models for Activity Recognition in Smart Homes

John W Kasubi[1*], Manjaiah D Huchaiah[2] , Mohammad Kazim Hooshmand[3]

[1].Local Government Training Institute Tanzania
[2].Department of Computer Science, Mangalore University, India
[3].Department of Computer Science, Kabul Education University, Afghanistan

Received: 15 Mar 2022/ Revised: 04 Aug 2022/ Accepted: 26 Sep 2022

Abstract

Activity Recognition is essential for exploring human activities in smart homes in the presence of multiple sensors as residents interact with household appliances. Smart homes use intelligent IoT devices linked to residents' homes to track human behavior as humans interact with the home's equipment, which may improve healthcare and security issues for the residents. Although remarkable studies have been done for pattern recognition and prediction of human activities in smart homes based on single residents and multiple residents using wearable sensors. However, not much research has been done on using Activity Recognizing Ambient Sensing (ARAS) residents. In this paper, we suggested using the ARAS dataset and newly emerged algorithms such as Deep learning Models to predict the activities of daily living (ADL). We compared the performance of deep learning models (ANN, CNN, and RNN) with that of classification models (DT, LDA, Adaboost, GB, XGBoost, MPL, and KNN) to figure out the ADL in the smart home residents. The experimental results demonstrated that DL models outperformed with an excellent accuracy compared to conventional classifiers in houses A and B in recognizing ADL in smart homes. This work proves that Deep Learning Models perform best in analyzing ARAS datasets compared to traditional machine learning algorithms.

Keywords: Conventional Classifiers; Deep Learning Model; Activity Recognition; Smart Homes; IoT; Feature Selection.

1- Introduction

The recognition of activities contributes to the improvement of multi-quality residents and security in a smart home environment by recognizing their activities of daily living (ADL) through using both Machine Learning (ML) Algorithms and Deep Learning (DL). Due to many tragedies happening abruptly in human life, such as covid-19, many tragedies have created a need for people to take care of their health; Smart Homes have become a solution [1]. Activity identification is critical in identifying and monitoring ADL in Smart Homes, resulting in a better life for residents of smart homes. The study was carried out using the Activity Recognition with Ambient Sensing (ARAS), collected from two houses named houses A and B, using the installed sensor of different household appliances, which involved 27 various activities. This study used DL Models and popular Conventional Classifiers, i.e., DT, LDA, Adaboost, GB, XGBoost, MPL, and KNN. DL is one of the key players in facilitating data analytics and learning in the IoT field and gives more accurate results and stable predictions.

Deep Learning (DL) is an algorithm that imitates the activities of the human brain to identify associations among massive amounts of data. DL technique learns complex functions and maps input to output directly from data by automatically learning features at multiple levels of abstraction. It is used to create algorithms to predict complex patterns and problems. It can adapt to changing inputs, allowing the network to produce the best possible result without redesigning the output criteria. DL is smart enough to learn and map nonlinear and complex relations, which is essential because many of the relationship issues between actual inputs and outputs are nonlinear and complex. After gaining knowledge from the preliminary information and their interrelations, DL can assume things on the unforeseen relationship issues with testing data, allowing the model to draw conclusions and predict the

✉ **John W Kasubi**
John Kasubi7@gmail.com

testing data. DL differs from many other prediction methods in that it does not impose limits on the input values; additionally, several experiments have shown that DL can model better and produce better results [2, 3].
On the other hand, the selected popular Conventional Classifiers are used to create robust models in classification problems [4, 5]. The Conventional Classifier techniques have been applied successfully in human activity recognition and achieved a reasonable recognition rate after feature selection and extraction on the ARAS dataset. However, few studies were conducted to compare conventional and deep learning in classifying human activities in the multiresident environment of smart homes.

The Motivation and contributions of this paper sought to fill the void in smart homes by developing a robust model capable of extracting hidden information and insights to improve prediction accuracy by applying newly emerging techniques. The study contributes to the research community of human activity recognition in smart homes to improve different aspects of human lifestyle such as health status, security and safety, monitoring and controlling energy and water usage, reducing living expenses, and thus improving quality of life. The better the model, the better the quality of life for smart home residents, is reducing expenditures on various items at home such as electricity, and water, increasing healthcare and security for residents.

This paper is structured as follows: The second Part briefly describes the interrelated works on Conventional Classifiers and Deep Learning; Third Section, presents Research Methodology used in this research; Fourth Part presents the analysis, performance, and discussions; and Fifth Part concludes and makes recommendations for future work.

2- Literature Review

This section explains previous related works reviewed concerning activity recognition for multiresident in smart homes, and the reviewed related research are as follows:

Natani et al. [6] demonstrated human activity recognition using the ARAS dataset to identify ADL. In this study, two types of RNN, GRU and LSTM, were used to simulate the various activities of the multiple residents in House A. The outcomes for 10days of GRU obtained 76.57% accuracy. In comparison, LSTM achieved 74.82% accuracy, for 30days, GRU obtained 80.35% accuracy while LSTM reached 78% accuracy, and finally, for 50days, GRU obtained 80.5% accuracy while LSTM achieved 79.08% accuracy, while on average, the author obtained 78%.

Bhattacharjee et al. [7] studied human activity classification to recognize different human activities using PNN, SVM, BPNN, and RNN techniques. The study identified other ADLs, and the experimental results were 94.10%, 59.11%, 97.40%, and 97.55% accuracy, respectively. The RNN outperformed the rest of the model by achieving 97.55% accuracy, predicting ADL.
Wang et al. [8] researched activity recognition using a deep learning algorithm based on the sensor. The authors suggested Deep Learning Algorithms used in identifying activities of daily living (ADLs) in smart homes because they have been proved to give better accuracy in model prediction.
Liciotti et al. [9] proposed using DL applications to identify human activities in Home Automation. An experimental outcome indicates that the LSTM method outperforms the existing DL and ML methods, producing better results than the current literature. The authors suggest more research to test other similar data sets for comparative analysis on activity detection.
Polat [10] developed a deep learning model to extract input data features automatically. In this regard, the researchers used LSTM, CNN, DBN, and RNN to test and train the models. The outcomes show that the suggested DL obtained an accuracy of 82.41%. The researcher recommends different human activity datasets and deep learning models and classifiers to enhance the model's efficiency.
Vakili et al. [11] compared eleven ML methods and DL for classification problems using six datasets. The comparison was conducted using different performance evaluation metrics. The experimental results show that RF performed better than other classifiers while ANN and CNN outperformed DL models.
Alshammari et al. [12] evaluated the performance of machine learning methods for ADL in Smart Homes; for this matter, the researchers employed several classifiers: DT, SVM, HMM, MPL, and Adaboost to address the problem. The experimental results demonstrate that the NN approach outperforms the other machine learning methods.
Tran et al. [13] proposed using edge intelligence in recognizing human activity in Smart Homes. For this case, they used both ML and DL algorithms to address the problem. Thus, CNN and SVM were adopted for activity recognition. The experiments were done, and DL outperformed ML techniques; the model achieved an accuracy of 95% in activity recognition. The authors suggested that other neural models be investigated in future work to improve the accuracy.
Park et al. [14] employed several deep neural networks to analyze residents' activities in a smart home using the MIT dataset. The experimental findings demonstrate that LSTM and GRU outshone other DL models; however, the dataset was too small to determine the best accuracy.

Akour et al. [15] performed a comparative study between standard traditional classifiers and deep learning to address ADL's effectiveness for older people. The CNN provided promising results in predicting ADL compared to ordinary conventional machine classifiers.

Igwe et al. [16] established a supervised learning algorithm known as a margin setting algorithm (MSA). They used ARAS as a data set to recognize patterns in the activity of daily living ADL) for both two residents in the smart home. Researchers obtained an average activity accuracy of 68.85% for house A and 96.24% for house B from the experiments. Despite the models outperforming researcher suggested conducting a comparative study between supervised learning algorithms with other different ML classifiers in a larger dataset scale.

Yun et al. [17] conducted a comparative analysis between classical machine classifiers (RF, SVM, IBL, and BayesNet). Deep learning algorithms were performed to detect human movements in smart homes using accuracy, precision, and recall evaluation metrics. Deep Learning outperformed with an accuracy of 90% compared to classical machine classifiers, which demonstrated poor performance.

3- Methodology

This section explains the methods deployed in this study, including the selected classifiers and the architecture of the activity recognition method.

3-1- Deep Learning Algorithm

Deep Learning (DL) was deployed in this study to identify human activities in smart homes using the ARAS dataset. The DL is a powerful NN formed by sophisticated mathematical modeling of various hidden layers in the NN and analyzing the data in a complex manner. In IoT data analytics, the DL Model is the most successful, produces the best results, and has been better than the conventional classifier [18, 19].

The DL is the most powerful among ML algorithms that process the input data to extract hidden insights from the dataset using dense layers, improving model accuracy. DL trains use massive amounts of data, eliminating the need to do a feature extraction manual as per conventional classifiers. Figure1 shows the Deep learning architecture model whereby the input layers receive binary data from observations. The binary data must be normalized or standardized to minimize the model's error and achieve the best model accuracy. The hidden layers use mathematical calculations on input data and nonlinear processing units to extract and transform features, while the output layers produce the desired results [20, 21]

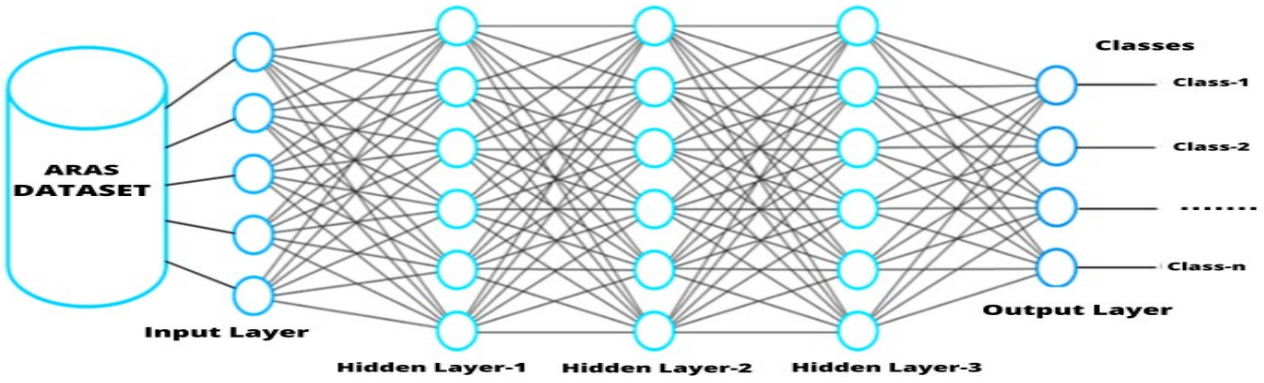


Fig.1 Structure of Deep Learning Model

1-1- Figure 2 shows the architecture of the Activation Function with inputs (X1, X2, X3, X4, + …….Xn), where f(s) is a nonlinear function known as the activation function Oj as an output value of the current neuron. The primary role of the Activation Function is that it is used to calculate and decide the output of a neural network.

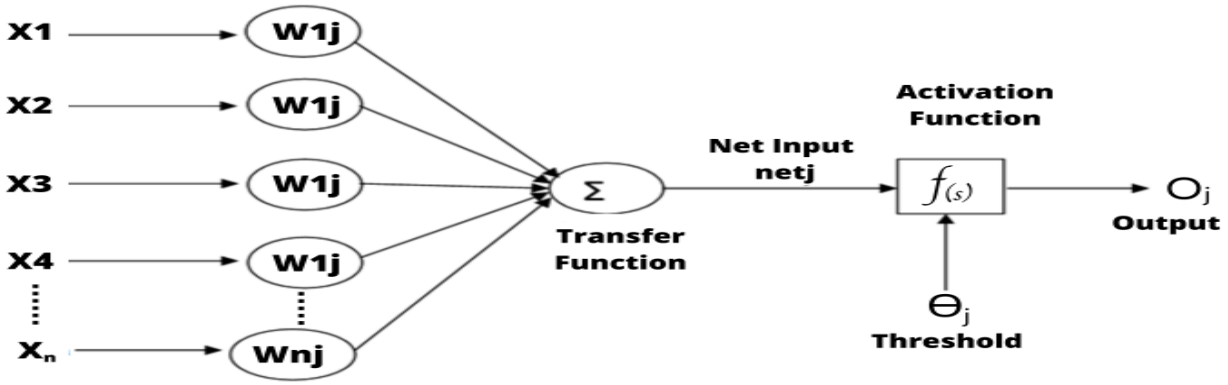


Fig. 2 Activation functions in neural networks

Figure 3 shows the suggested architecture for human activity recognition in multiresident based on smart homes using Deep learning and the ARAS dataset [22]. Before using Deep Learning to train the model, the ARAS data preprocessing was used to clean the dataset, perform feature scaling, and compute the sample size. Feature scaling was utilized to reduce model complexity while also increasing model accuracy. To ensure that we managed to achieve our goal, we used MinMaxScaler to sparse the datasets into zeros (0) and ones (1)

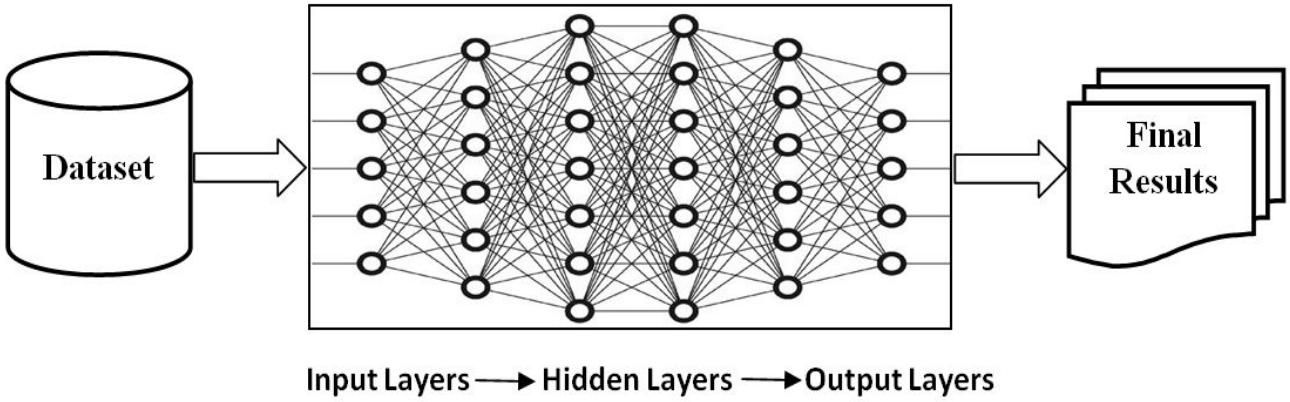


Fig.3 Proposed Deep Learning Architecture

3-2- Conventional Classifiers

On the other hand, the study employed several popular conventional classifiers to recognize human activities in the Smart Homes using the ARAS dataset. This study employed a popular conventional classifier for comparative research with DL, which includes; DT-is the method used for solving classification problems, which uses internal nodes to represent a predictor variable. In this tree-structured classifier, each leaf node represents the outcome of the majority voting, and it is applied to and utilized in more than one classification [23]. LDA - is a statistical technique for binary and multiclass classification that reduces the number of features to a more manageable number before classification by assigning objects to one group among several groups. Hence, increasing the model accuracy [24]. Adaboost is a method used as an ensemble technique to build multiple models of the type using a sequential set of algorithms, reduce bias and variance, and convert weak learners into strong ones to create a robust model to improve the performance [25]. Gradient boosting (GB) is an ensemble strategy for enhancing the model's prediction performance using ensembles. Decision trees are often used because they combine multiple weak classifier models to create a robust predictive model employing a set of classifiers [26]. XGBoost method is a type of ensemble method that uses the framework of gradient boosted the decision tree to tackle classification

tasks. It uses enhanced regularization (L1 & L2) and parallel computation [27]. Multi-layer perceptrons (MLP) are often used for training input-output pairs for problem classifying and predicting input-output relationships. Training entails fine-tuning model parameters to reduce errors and thus improve model performance [28]. KNN- is the most straightforward algorithm used to classify a new data point into a target class depending on the features of its neighboring data points. The KNN algorithm believes that identical items are close to each other, and for better accuracy, it uses turning parameters to select the correct value of 'k' [29].

The reasons for selecting the above-mentioned conventional classifiers are suitable for the multiclassification problem, simple to implement, fast to train and overcome overfitting, ability to compress the dataset into a manageable size, and ability to produce a robust model. Activity recognition plays a vital role in Smart homes by maintaining the residents' well-being and making life more meaningful. It helps enhance multiresidents quality of life and health in a smart home neighborhood

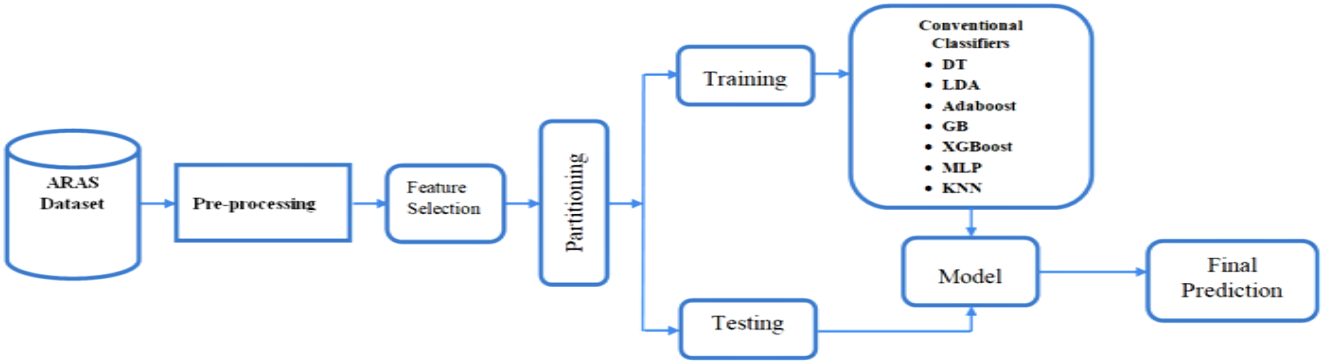


Fig.4 Proposed Approach for conventional classifiers

3-3- Data Preprocessing

Data preprocessing is a data mining technique that transforms raw data into an understandable format. For this reason, we employed feature selection, and feature scaling. We calculated the sample size before creating models using DL and conventional classifiers in multiclass classification problems using the ARAS dataset.

3-3-1 Feature Selection

Feature Selection: A secret to the performance of any algorithm is the selection of relevant features; removing irrelevant features in the dataset reduces the computing complexity of the model, which in turn leads to outstanding accuracy. Feature selection was done to minimize overfitting, speed up training time, and improve the model accuracy. Univariate feature selection was employed to select randomly the 10 best features that have a strong relationship with the target variables. For this matter, we employed the sklearn library that provides the SelectKBest class that uses the chi-squared (chi2) statistical test to select the 10 best features from the ARAS dataset that are strongly dependent on the response [30].

$$X_C^{\,2} = \sum \frac{(O_i - E_i)}{E_i} \quad (1)$$

Where; c – is the degree of freedom; O – is the observed value(s) and E–is the expected value(s)

3-3-2 Feature Scaling

Feature scaling was used to scale all values into the range of 0 and 1 to reduce model complexity and increase the model's accuracy. It was carried out using MinMaxScaler to sparse the datasets into zeros (0) and ones (1) to make sure that we achieve the best accuracy with the selected Conventional Classifiers and DL [31].

$$X = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

Where; X – is the normalized data, X_i – is the original feature value, X_{min} – is the minimum value, and X_{max} – is the maximum value in the original dataset before scaling.

3-3-3 Imbalanced Dataset

The ARAS dataset is imbalanced, so the SMOTE technique was applied to balance the dataset to solve this problem. Then, the dataset was divided into training and testing sets; for this reason, the imbalanced Learn library that provides the imblearn class was applied to cater to the imbalanced problem. After that, models were built using Conventional Classifiers and DL (DT, LDA, Adaboost, GB, XGBoost, MPL, KNN, and DL). Hence, a comparison between Deep learning and Conventional Classifiers was performed; DL was outshone compared by Conventional Classifiers [32, 33].

4- Experimental Results and Discussions

This section explains the ARAS dataset, the findings, and discussions of the suggested methods for activity detection in multiresidents based on the ARAS dataset's smart homes. This study experimented with both DL and Conventional Classifiers using the ARAS dataset.

4-1-Experimental Setup

The data used during this research was collected by the ARAS (Activity Recognition with Ambient Sensing) dataset for multiresidents in smart homes to detect activity. The ARAS dataset was collected from two different real houses for two months in Turkey in 2013. The dataset involved 27 different types of activities and contained a total of 5,184,000 instances from each house which is a large dataset [34]. In this regard, both conventional classifiers and Deep Learning were employed to draw significant insight from Activities of Daily Living (ADL).

4-2- Evaluation Matrices

The study used four evaluation methods to examine the performance of our model, including Classification Accuracy (CA), recall, precision, and F1-measure. These metrics were used to evaluate the model's performance because accuracy alone is not enough to infer a model's performance.

Accuracy: Is the value of the forecast divided by the total forecasting value

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3}$$

Precision: Is the actual positive value divided by the positive class value and false positive value.

$$\text{Pr}ecision = \frac{(TP)}{(TP + FP)} \tag{4}$$

Recall: It is called the True Positive rate. The positive truth value is divided by the actual positive and false negative values

$$\text{Re}call \frac{(TP)}{(TP + FN)} \tag{5}$$

F-1 Measure: Mean of Precision and Recall

$$F1 - Score = 2 * \frac{(\text{Pr}ecision * \text{Re}call)}{(\text{Pr}ecision + \text{Re}call)} \tag{6}$$

Whereas TP represents True Positive values, TN is a True Negatives value, FP is a False Positive, and FN is a False Negatives value.

4-3- Analysis

This section provides a comparative analysis obtained while implementing the proposed approach in developing a predictive model for activity recognition in multiresidents in a smart home environment using both Deep Learning and conventional classifiers. We first loaded the ARAS dataset and then loaded the basic libraries; we created the sequence model with dense layers. First, we constructed the dense layer with 128 neurons, and like the first, we had to specify the number of input dimensions (20), and ReLU was used as an activation function, the next layer was the dense layer with 256 neurons, and ReLU was used as an activation function; then a dropout layer with 0.2% as the techniques used to overcome the issue of overfitting during the training of the model. After that, we had another dense layer with 64 neurons, and ReLU was used as an activation function. Finally, we had a dense output layer, and softmax was used as an activation function; it converts the results in probability values. Next, we compiled the model, and since this is a multiclass classification, we used categorical cross entropy as the loss function and softmax as an optimizer. We also used categorical accuracy as a metric. Next, we trained the model using epochs=300 and batch_size=128; after that, we evaluated our model using a test dataset, and the model achieved an excellent accuracy compared to the conventional classifier. Finally, we cross checked the correctness of the predicted and expected values using the loop function and plotted model accuracy and model loss curves.

On the other hand, conventional classifiers: The models were created using the sklearn library; in the preprocessing data stage, we applied feature scaling in the input values before developing a model for predictions to reduce the scatteredness of the data. For this matter, we used MinMaxScaler to carter for feature scaling in conventional classifiers. The ARAS dataset was divided into training and testing sets; then, models were developed using DT,

LDA, Adaboost, GB, XGBoost, MLP, and KNN. The performance metrics such as accuracy, precision, recall, f1-score, and correlation matrix were applied to evaluate the performance of the model. Hence, the model prediction was done to cross-check the correctness of the predicted and expected values using the loop function.

4-4- Findings and Discussion

This part describes the findings of the experimental tests and discussions for multiresident activity detection in a smart home using both the Deep learning (DL) method and seven conventional classifiers (DT, LDA, Adaboost, GB, XGBoost, MLP, and KNN) together with performance metrics such as accuracy, precision, recall, f1-score and correlation matrix. The results show that DL outshone seven conventional classifiers in both houses A and B for activity identification for multiresidents. Furthermore, DL performed best in house B compared to house A, and conventional classifiers performed best in house B compared to house A. Table 1 and Table 2 show the outcomes achieved by Deep learning compared to the seven conventional classifiers used in this study aligned with the discussion.

Table 1: Classification performance comparison in House A

Classifiers	Evaluation Metrics			
	Acc (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	0.6999	0.685	0.683	0.667
LDA	0.6349	0.634	0.624	0.596
Adaboost	0.5951	0.562	0.567	0.515
GB	0.6960	0.673	0.655	0.594
XGBoost	0.6969	0.685	0.687	0.634
MLP	0.6945	0.687	0.696	0.603
KNN	0.6921	0.676	0.675	0.684
ANN	0.9944	1.00	0.993	0.993
CNN	0.9916	0.9921	0.991	0.993
RNN	0.9898	0.989	0.989	0.989

As shown in Table 1, the experimental results for both conventional classifiers and Deep learning models regarding the accuracy, precision, recall, and F1 score. The deep learning model outscored with precision accuracy of 100% compared to conventional classifiers, which performed moderately. The conventional classifier's performance was DT 69.99% accuracy, followed by MLP - 69.45%, XGBoost-69.69%, GB-69.60%, KNN-69.21%, and LDA-63.49%, while Adaboost performed moderately compared to the rest classifiers with an accuracy of 59.51%. In addition, the results from Table 1 are demonstrated in figure 5 below.

Table 2: Classification performance comparison in House B

Classifiers	Evaluation Metrics			
	Acc (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	0.9193	0.912	0.935	0.914
LDA	0.8343	0.810	0.836	0.815
Adaboost	0.9036	0.898	0.903	0.905
GB	0.9084	0.914	0.921	0.913
XGBoost	0.9147	0.913	0.902	0.925
MLP	0.9113	0.924	0.913	0.905
KNN	0.9034	0.905	0.923	0.914
ANN	0.9983	1.00	1.00	1.00
CNN	0.9963	0.9963	0.995	0.995
RNN	0.9965	0.9965	0.997	0.996

Table 2 displays the experimental comparison results in house B for both conventional classifiers and deep learning models regarding the accuracy, precision, recall, and F1 score. The DL model outshone conventional classifiers in every measure with the precision, recall, and f1-score of 100% for activity recognition in house B. DT, XGBoost, and MLP came in second, with an accuracy of 91.93%, 91.47%, and 91.13%, respectively, approximately 6% lower than Deep learning. However, when compared to the other classifiers, LDA scored less, with an accuracy of 83.43%.

4-5- Comparative Analysis

Table 3 demonstrates the comparison between the prior study and the proposed approach. This study outperformed the previous studies in activity recognition using ANN by achieving an average precision, recall, and f1-score of 100%. In comparison, the earlier research by Natani et al. [6] achieved an accuracy of 81.7%, 79.25%, 70.9%, 83.61%, and 85.94%, 88.75%,90.85%, 88.87% in houses A and B, by using RNN, CNN, MLP, and GRU, respectively. Tran et al. [13] achieved 95% in house B using CNN, while Igwe et al. [16] obtained 67.32%, 68.85%, and 67.32%, 68.85% accuracy by using ANN and MSA. As a result, the proposed approach outperformed the earlier experiments in activity recognition by achieving an accuracy of 99.4%, 99.16%, 98.98%, 69.45%, and 99.83%, 99.63%, 99.65%, 91.132%

in houses A and B, respectively using ANN, CNN, RNN, and MLP.

Table 3: Classification with Previous Research

Research Study	Method	Accuracy House A	Accuracy House B
Natani et al. [6]	ANN, MSA	67.32%, 68.85%	95.43%, 96.24%
Tran et al. [13]	CNN	-	95%
Igwe, et al. [16]	RNN, CNN , MLP, GRU	81.7%, 79.25%, 83.61%, 70.9%	85.94%, 88.75%, 88.87%, 90.85%
The proposed approach	**ANN, CNN, RNN, MLP**	**99.4%, 99.16%, 98.98%, 69.45%**	**99.83%, 99.63%, 99.65%, 91.132%**

5- Conclusions and Future Directions

This study presents a novel comparative analysis study between conventional classifiers and deep learning (DL) models. The experimental results show that Deep learning models outperformed in both houses A and B compared to conventional classifiers. The ANN outperformed other DL models and all ML classifiers with an average score of 100% for precision, recall, and f1-score in house B; in predicting human activities using the ARAS dataset. However, conventional classifiers performed best in house B compared to house A. The experimental results prove that the Deep learning methods analyze ARAS datasets better than conventional classifiers. In comparison between the prior study and the proposed approach, this study outperformed the previous studies in activity recognition using ANN by achieving an average precision, recall, and f1-score of 100%.

In future work, we suggest that different traditional machine learning classifiers to be employed on the ARAS dataset compared with Deep learning models.

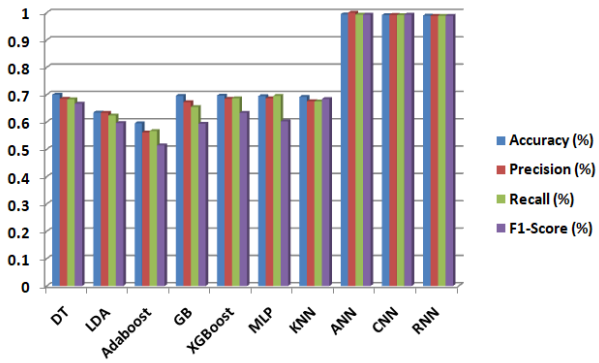


Fig. 5 Classification comparison in house A

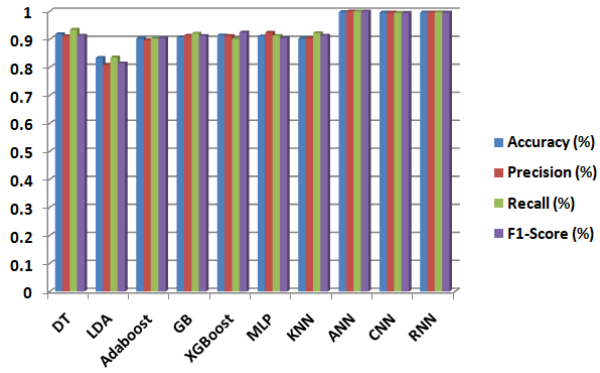


Fig. 6 Classification Comparison in House B

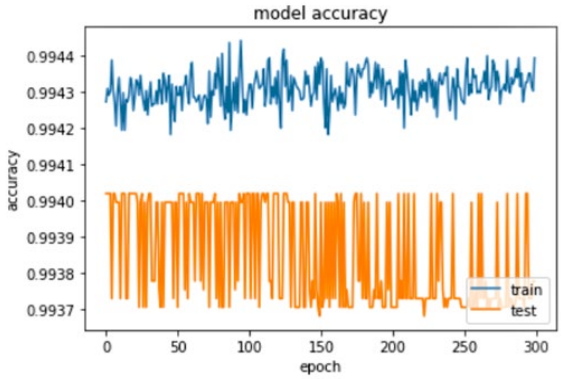


Fig. 7 Model Accuracy in House A using ANN

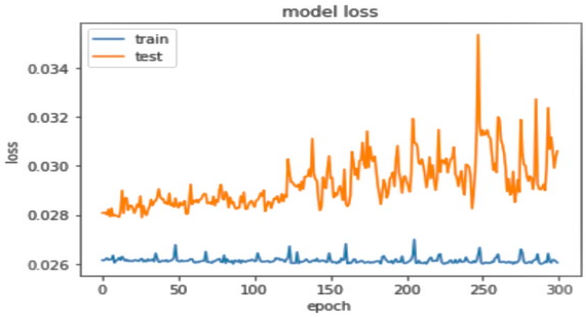


Fig.8 Model Loss in House A using ANN

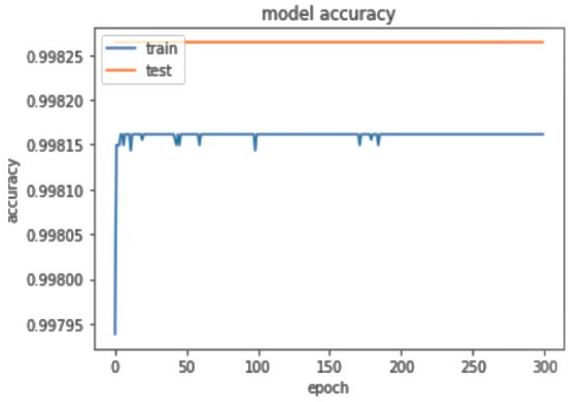


Fig.9. Model Accuracy in House B using ANN

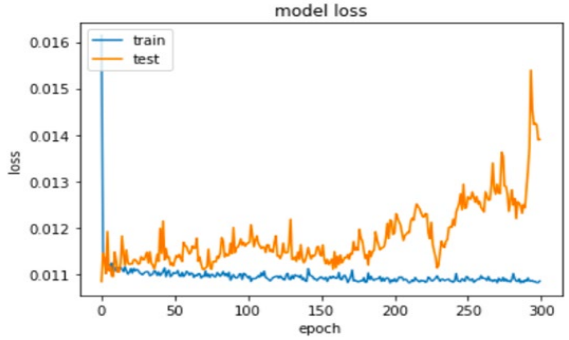


Fig. 10 Model Loss in House B using ANN

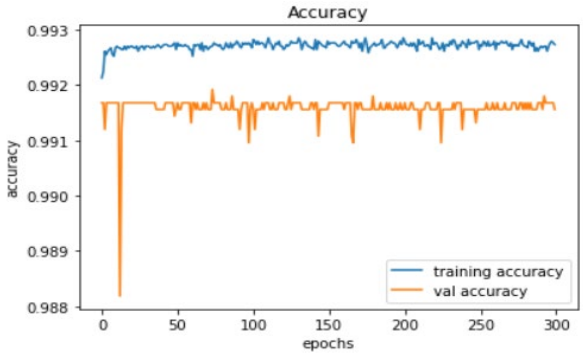


Fig.11 Model Accuracy in house A using CNN

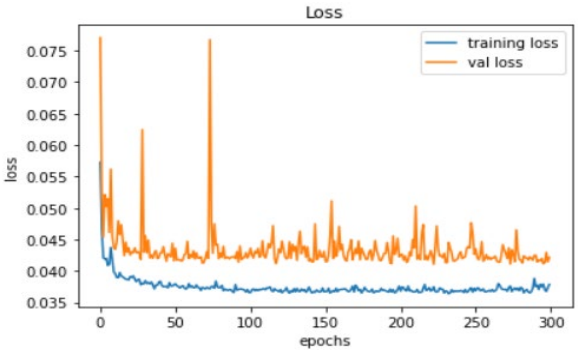


Fig. 12 Model Loss in House A using CNN

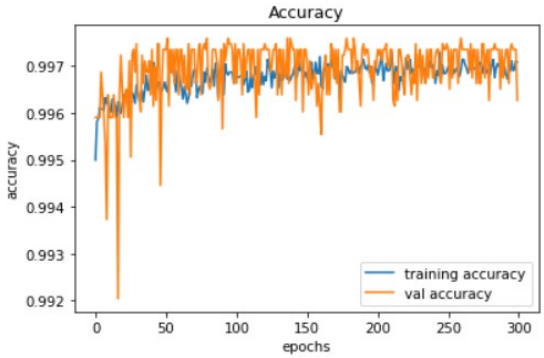


Fig. 13 Model Accuracy in House B using CNN

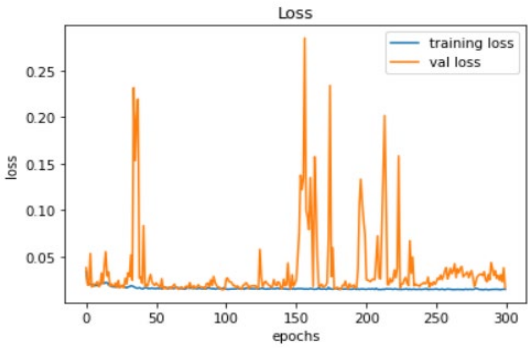


Fig. 14 Model Loss in House B using CNN

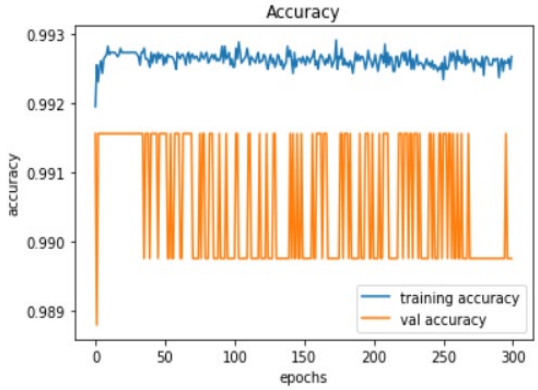


Fig. 15 Model Accuracy in House A using RNN

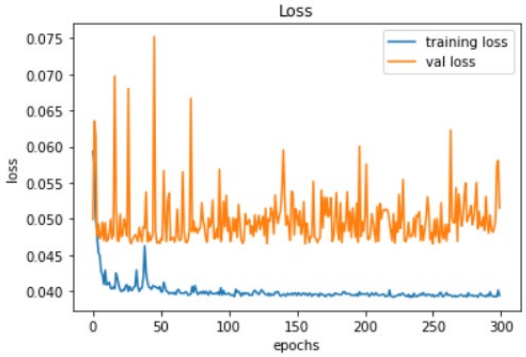


Fig.16 Model Loss in House A using RNN

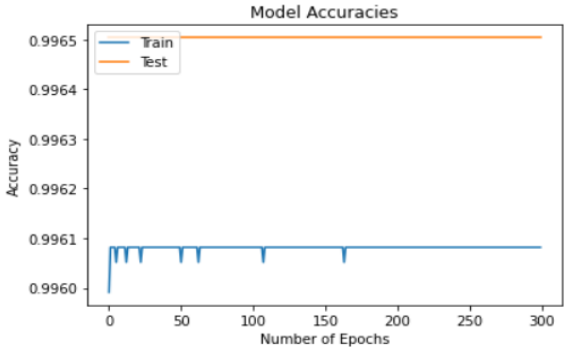


Fig. 17 Model Accuracy in House B using RNN

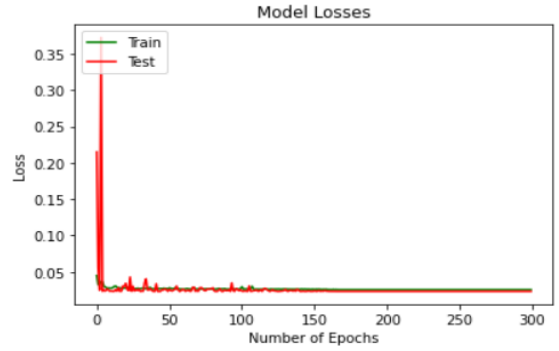


Fig. 18 Model Loss in House B using RNN

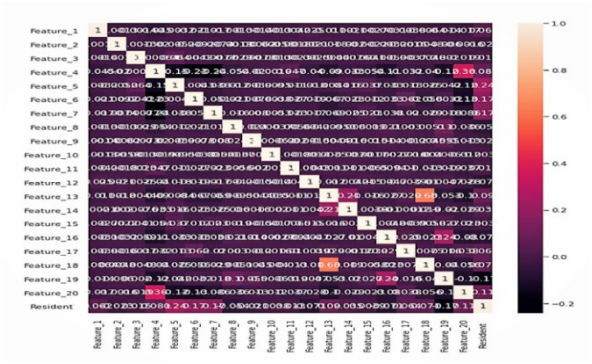


Fig. 19 Correlation Matrix with Heatmap for House A

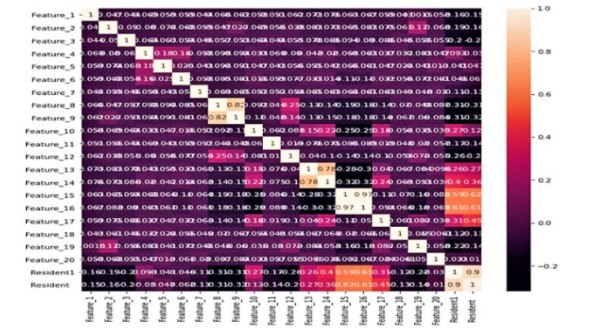


Fig. 20 Correlation Matrix with Heatmap for House B

References

[1] D, Emiro. P Ariza-Colpas, J. M. Quero, and M. Espinilla. "Sensor-based datasets for human activity recognition–a systematic review of literature." *IEEE Access* 6, 2018, pp.59192-59210.

[2] M. Mehdi, A. A. Sameh-Sorour, and M. Guizani. "Deep learning for IoT big data and streaming analytics: A survey." *IEEE Communications Surveys & Tutorials* 20, no. 4, 2018, pp.2923-2960.

[3] H. M. Mehedi, z. Uddin, A. Mohamed, and A. Almogren. "A robust human activity recognition system using smartphone sensors and deep learning." *Future Generation Computer Systems* 81, 2018, pp.307-313.

[4] F. Zhice, Y. Wang, L. Peng, and H. Hong. "Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping." *Computers & Geosciences* 139, 2020.

[5] J. Monika, N. Kesswani, and M. Kumar. "A Deep Learning Approach for Classification and Diagnosis of Parkinson's Disease.", 2021.

[6] N., Anubhav, A. Sharma, T. Peruma, and S. Sukhavasi. "Deep learning for multi-resident activity recognition in ambient sensing smart homes." In *2019 IEEE 8th Global conference on consumer electronics (GCCE),* 2019, pp. 340-341.

[7] B. Sayandeep, S. Kishore, and A. Swetapadma. "A comparative study of supervised learning techniques for human activity monitoring using smart sensors." In *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, 2018, pp. 1-4.

[8] W. Jindong, Y. Chen, S. Hao, X. Peng, and L. Hu. "Deep learning for sensor-based activity recognition: A survey." *Pattern Recognition Letters* 119, 2019, pp.3-11.

[9] L. Daniele, M. Bernardini, L. Romeo, and E. Frontoni. "A sequential deep learning application for recognising human activities in smart homes." *Neurocomputing* 396, 2020, pp.501-513.

[10] M. Homay D. and H. Polat. "Human activity recognition in smart home with deep learning approach." In *2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*, 2019, pp. 149-153.

[11] V. Meysam, M. Ghamsari, and M. Rezaei. "Performance analysis and comparison of machine and deep learning algorithms for iot data classification." *arXiv preprint arXiv:2001.09636* , 2020).

[12] A. Talal, N. Alshammari, M. Sedky, and C. Howard. "Evaluating machine learning techniques for activity classification in smart home environments." *Int. J. Inf. Commun. Eng* 12, 2018, pp.72-78.

[13] T. Son N., D. Nguyen, T. Ngo, X. Vu, L. Hoang, Q. Zhang, and M. Karunanithi. "On multi-resident activity recognition in ambient smart-homes." *Artificial Intelligence Review* 53, no. 6, 2020, pp.3929-3945.

[14] P. Jiho, K. Jang, and S. Yang. "Deep neural networks for activity recognition with multi-sensor data in a smart home." In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, 2018, pp. 155-160.

[15] M. Akour, O. Al Qasem, H. Al Sghaier, and K. Al-Radaideh. "The effectiveness of using deep learning algorithms in predicting daily activities." *International Journal* 8, no. 5, 2019.

[16] I. Ogbonna M. Y. Wang, G. C. Giakos, and J.Fu. "Human activity recognition in smart environments employing margin setting

algorithm." *Journal of Ambient Intelligence and Humanized Computing*, 2020, pp. 1-13.

[17] Y. Jaeseok, and J. Woo. "A comparative analysis of deep learning and machine learning on detecting movement directions using PIR sensors." *IEEE Internet of Things Journal* 7, no. 4, 2019, pp.2855-2868.

[18] D.G. Reza, X. Chen, and W. Yang. "A Review of Artificial Intelligence's Neural Networks (Deep Learning) Applications in Medical Diagnosis and Prediction." *IT Professional* 23, no. 3, 2021, pp.58-62.

[19] A. Rai, H. Md Junayed, Z. Ahmad, and J. Kim. "A Fault Diagnosis Framework for Centrifugal Pumps by Scalogram-Based Imaging and Deep Learning." IEEE Access 9, 2021, pp.58052-58066.

[20] D. Shon, H. Md Junayed, K. Im, H. Choi, D. Yoo, and J. Kim. "Sleep state classification using power spectral density and residual neural network with multichannel EEG signals." Applied Sciences 10, no. 21, 2020.

[21] M. Sohaib, H. Md Junayed, and J. Kim. "An explainable ai-based fault diagnosis model for bearings." Sensors 21, no. 12, 2021.

[22] C. Kaixuan, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu. "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities." *ACM Computing Surveys (CSUR)* 54, no. 4, 2021, pp.1-40.

[23] L. Sidrah, K. Dashtipour, S. A. Shah, A. Rizwan, A. A. Alotaibi, T. Althobaiti, K. Arshad, K. Assaleh, and N. Ramzan. "Novel Ensemble Algorithm for Multiple Activity Recognition in Elderly People Exploiting Ubiquitous Sensing Devices." *IEEE Sensors Journal* 21, no. 16, 202, pp.18214-18221.

[24] S.Pekka, and J. Röning. "Context-aware incremental learning-based method for personalized human activity recognition." *Journal of Ambient Intelligence and Humanized Computing* 12, no. 12, 2021, pp.10499-10513.

[25] T. Pratik, and I. Bose. "Recognition of human activities for wellness management using a smartphone and a smartwatch: A boosting approach." *Decision Support Systems* 140, 2021.

[26] S. Ryoichi, K. Abe, T. Yokoyama, M. Kumano, and M. Kawakatsu. "Ensemble learning for human activity recognition." In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, 2020, pp. 335-339.

[27] V. Indumathi, and S. Prabakeran. "A Comparative Analysis on Sensor-Based Human Activity Recognition Using Various Deep Learning Techniques." In *Computer Networks, Big Data and IoT*, Springer, Singapore, 2021, pp. 919-938.

[28] J. Qiang, S. Guo, P. Chen, P. Wu, and G. Cui. "A Robust Real-time Human Activity Recognition method Based on Attention-Augmented GRU." In *2021 IEEE Radar Conference (RadarConf21)*, IEEE, 2021, pp. 1-5.

[29] B. M. Hashim, K. Mohammed, and R. Amutha. "Machine Learning-based Human Activity Recognition using Neighbourhood Component Analysis." In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1080-1084.

[30] L. Yaqing, Y. Mu, K. Chen, Y. Li, and J. Guo. "Daily activity feature selection in smart homes based on pearson correlation coefficient." *Neural Processing Letters* 51, no. 2 , 2020, pp.1771-1787.

[31] M. Mojtaba, and N. Wilson. "Scaling-invariant maximum margin preference learning." *International Journal of Approximate Reasoning* 128, 2021, pp.69-101.

[32] H. Rebeen A. M. Kimura, and J. Lundström. "Efficacy of imbalanced data handling methods on deep learning for smart homes environments." *SN Computer Science* 1, no. 4, 2020, pp.1-10.

[33] H.Md Junayed, Jia Uddin, and Subroto Nag Pinku. "A novel modified SFTA approach for feature extraction." In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2016, pp. 1-5.

A. Hande, and C. Ersoy. "Multi-resident activity tracking and recognition in smart environments." *Journal of Ambient Intelligence and Humanized Computing* 8, no. 4, 2017, pp.513-529.

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 138-151
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Designing a Semi-Intelligent Crawler for Creating a Persian Question Answering Corpus Called Popfa

Hadi Sharifian[1] , Nasim Tohidi[2] , Chitra Dadkhah[2*]

[1].K. N. Toosi University of Technology, e-learning center, Tehran, Iran
[2].Artificial Intelligence Department, Computer Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran

Received: 28 Jan 2023/ Revised: 04 Oct 2023/ Accepted: 12 Nov 2023

Abstract

Question answering in natural language processing is an interesting field for researchers to examine their ability in solving the tough Alan Turing test. Everyday computer scientists are trying hard to develop and promote question answering systems in various natural languages, especially English. However, in Persian, it is not easy to advance these systems. The main problem is related to low resources and not enough corpora in this language. Thus, in this paper, a Persian question answering text corpus is created, which covers a wide range of religious, midwifery, and issues related to youth marriage topics and question types commonly encountered in Persian language usage. In this regard, the most important challenge was introducing a method for data gathering in Persian as well as facilitating and expanding the data gathering process. Though, SIC (Semi-Intelligent Crawler) is proposed as a solution that can overcome the challenge and find a way to crawl the Persian websites, gather text and finally import it to a database. The outcome of this research is a corpus called Popfa, which stands for Porsesh Pasokh (question answering) in Farsi. This corpus contains more than 53,000 standard questions and answers. Besides, it has been evaluated with standard approaches. All the questions in Popfa are answered by specialists in two general topics: religious and medical questions. Therefore, researchers can now use this corpus for doing research on Persian question answering.

Keywords: Question Answering; Persian Corpus; Religious Questions; Medical Questions; Natural Language Processing.

1- Introduction

It has been many years since Alan Turing introduced his famous experiment, and despite all the advances that have taken place in the world of computer science, no computer or even supercomputer has yet successfully passed the Turing test completely. This simple experiment is an artificial intelligence that communicates with a human through a computer user interface and convinces the human that he is communicating with a human. [1]

Designing and implementing an efficient question answering system, which can provide the accurate answer to the user input question in natural language in the shortest time, is one of the most attractive and practical problems in the field of artificial intelligence for computer scientists and researchers as well as managers of companies providing computer technology services such as the production of programs, websites, speech bots, etc. In fact, this question answering system can be a model of the same machine that is supposed to pass the Turing test successfully. There have been astonishing advances in English in this area with scientists achieving successful results in terms of the Turing experiment, but in Persian, despite several tools which have been proposed in recent years [2-6], research on question answering datasets has not progressed significantly [7]. One of the reasons for the abandonment of the ancient and rich Persian language in this category is the inadequacy of a comprehensive and powerful corpus of valid questions and answers. Currently, to the best of our knowledge, the largest Persian question answering corpus was far much smaller than the similar one for other languages like English.

Persian (Farsi) language has many attributes that make it distinct from other well-studied languages. In terms of script, Persian is similar to Semitic languages, like Arabic and Amharic. Linguistically, however, Persian is an Indo-European language [8,9] and thus distantly related to most of the languages of Europe as well as the northern part of

✉ **Chitra Dadkhah**
dadkhah@kntu.ac.ir

the Indian subcontinent. Therefore, the Persian language has features that distinguish it from the English language and make its processing more complex. For example, in Persian, some letters stick to each other, and in addition to the space between words, the half-space is also used in writing Persian text. In addition, the structure of sentences and the way in which words with different roles are placed in Persian sentences is not the same as in English sentences. Therefore, the methods introduced for English cannot be used for Persian. Another important point is that many of the texts and data available are not written in the formal language, for example, the half-space is not observed all the time. These are some of the reasons why the number of research done on the Persian language is very small compared to the English language.

Today, Google is known as an intelligent search engine. However, this powerful search engine returns many links for each incoming question that are not necessarily the intended result. Therefore, users must open links and check the content of each one to finally find the answer among a large number of returned links. This is where having a question answering system in Persian that receives a question and provides an accurate answer seems necessary. In this regard, the purpose of this paper is to create a question answering corpus in the Persian language.

The structure of the paper is as follows: In Section 2, some previous works in this field are introduced. Then, Section 3 explains the proposed method. Section 4 contains the steps for implementing our proposed strategy. Sections 5 and 6 describe evaluations and experimental results, respectively. More, section 7 presents the prepared user interface and section 8 gives a discussion about the unique feature of the corpus. Finally, in Section 9 conclusion and future works are summarized.

2- Related Works

The number of available systems for Persian language processing is very small compared to the English language, which has led to a decrease in the research on Persian language in the field of natural language processing [5,3]. There is a lack of standard systems in the field of Persian language question answering systems, as one of the applications of language processing.

Since 1999, TREC (Text Retrieval Conference) has had a question answering track [10] resulting in high accuracy systems for English, like methods in [11] and [12].

In some related papers, researchers have decided to either apply community-sourced datasets or develop restricted-domain question answering systems. For instance, in [5] the Rasekhoon[1] question answering dataset was used to evaluate a question matching model in Persian. Plus, in [13], TriviaQA was presented, which was a reading comprehension dataset including question-answer-evidence triples. It contained question-answer pairs written by trivia enthusiasts and gathered evidence documents (on average 6 per question) which provided distant supervision for answering the questions.

The main activities in the field of Persian question answering systems, like [7], [14], and [15], have focused on approaches based on the feature that the question raised in Persian can be analyzed from a syntactic and semantic perspective, and the most appropriate answer can be selected based on the available database.

In [10], authors introduced a standard Persian text collection, named Hamshahri, which was built from a large number of newspaper articles according to TREC specifications in which statistical information about documents, queries, and their relevance judgment were presented. This collection can be downloaded as a package from its website. The package contains all relevant judgments for the 65 standard topics, some descriptions of previous research conducted based on the collection, and some source codes for indexing and retrieval of the collection [16].

In [17], sentences were classified into two levels of coarse and fine classes based on the type of answer to each question. After extracting features and setting a sliding window on the Conditional Random Fields (CRF) model, CRF Question Classifier (QC) was trained to predict labels for every token in question. Then, a majority voting on the question classification output was used to extract a unique label for each question, and the effects of features on the ultimate accuracy of the system were evaluated.

Also, in [18], they proposed an approach that was used in an online automatic question answering system. They combined rule-based and machine learning question classification approaches for highly inflectional languages such as Persian. They got satisfactory results according to the high number of question classes.

In [19], a cross-lingual approach using a unified semantic space among languages was introduced. In this study, after keyword extraction, entity linking, and answer type detection, cross lingual semantic similarity was used to extract the answer from the knowledge base via relation selection and type matching.

In [20], a corpus for the Persian language was presented. This corpus consists of 2,118 non-factoid and 2,051 factoid questions and for each question, question text, question type, question difficulty from the questioner and responder perspective, expected answer type in coarse-grained and fine-grained level, the exact answer, and page and paragraph number of answer are annotated. This corpus can be applied to learn components of a question answering system, including question classification, information retrieval, and answer extraction. This corpus is freely available for academic purposes.

[1] www.rasekhoon.net

In [15], a medical question answering system for the Persian language was proposed. In their research, a dataset of diseases and drugs was collected and structured. The system included three main modules: question processing, document retrieval, and answer extraction. For the question processing module, a sequential architecture was designed which retrieved the main concept of a question by using different components. In these components, rule-based methods, natural language processing, and dictionary-based techniques were used. In the document retrieval module, the documents were indexed and searched using the Lucene library. The retrieved documents were ranked using similarity detection algorithms and the highest-ranked document was selected to be used by the answer extraction module. This module was responsible for extracting the most relevant section of the text in the retrieved document.

In [21], PeCoQ was defined which was a Persian question answering dataset. It included 10K questions and answers extracted from the Persian knowledge graph, FarsBase. Additionally, for each question, the SPARQL query and 2 paraphrases authored by linguists were provided. There were various complexity types in this dataset, like multi-relation, multi-entity, comparative, superlative, aggregation, ordinal, and temporal constraints.

In [22], a Persian Question Answering Dataset (ParSQuAD) was generated based on translating the SQuAD 2.0 dataset by machine. Through it, some errors have been detected within the process of translation; resulting in two different versions of it, depending on whether these errors have been corrected automatically or manually. The most important weakness of this dataset is that it does not have the quality of a native Persian reading comprehension dataset containing native question and answer samples annotated by multiple human annotators.

In [23], PersianQuAD was introduced which was a native question answering dataset for the Persian language. The authors built this dataset in 4 phases: 1) Wikipedia article selection, 2) question-answer collection, 3) three-candidate test set preparation, and 4) Data Quality Monitoring. The output dataset contained about 20K questions and answers made by native annotators on a set of Persian Wikipedia articles. The answer to each question was a segment of the corresponding article text. According to their report, PersianQuAD consisted of questions of different types and complexities. Plus, they proposed 3 versions of a deep learning-based question answering system trained using MBERT, ALBERT-FA, and ParsBERT on PersianQuAD, and for MBERT they achieved the best result.

Finally, in [24], authors proposed PQuAD, a crowdsourced reading comprehension dataset for Persian on Wikipedia articles which included various subjects. Its data collection process had 3 phases: 1) passage curation, 2) question-answer pair annotation, and 3) additional answer collection. The output dataset consisted of 80K questions and their answers. They evaluated different properties of the dataset to depict its diversity and complexity as a machine reading comprehension benchmark.

Considering all the mentioned efforts and information, proposing a feasible and accurate method for gathering questions and answers in Persian in a corpus to be used for training question answering systems in the future is crucial. The created corpus in this paper serves as a dataset that can be utilized to train and improve the performance of Persian question answering systems. These systems can leverage machine learning techniques, such as deep learning algorithms, to learn from the provided data and enhance their question answering abilities. Hence, in the following the proposed method for creating the corpus is explained.

3- The Proposed Method

Research conducted in the field of question answering systems shows the shortage of a standard Persian question answering corpus [7]. Naturally, scientists face various limitations and challenges in this area. Considering the incoherence of information related to Persian question answering, there could be two solutions:

- Solution 1: Generating basic questions and answers, followed by a Persian question answering system.
- Solution 2: Collecting questions and answers available on global websites in Persian and editing them to produce a Persian question answering corpus.

3-1- Challenges

As a creative and new way, a cost-benefit table was formed to select the logically desired solution. The most important selection features were:

1) Time taken
2) Research and executive costs
3) Required human resources
4) Technical and structural limitations
5) Verification capability
6) The size of the database
7) Domain comprehensiveness

Since these features were selected by a group without any previous background and are among the innovations of this paper, there was no reported quantitative data. Therefore, we decided to compare the features using two strategies of collection and production. In other words, what is the relationship between them regardless of the quantity of each one. The value comparison has been summarized as Low, Medium, and High cost.

The evaluation results of the above-mentioned features are shown in Table 1. Production strategy as well as fact-checking was preferred for technical and structural features, but it was significantly different from the collection strategy for other features. Of course, due to the pristine nature of this area, the execution time for either

solution was not clear, and we could only compare the time between the two solutions.
The scientific method of cost-benefit analysis in our proposed approach resulted in -10 for the collection method and -17 for the production method. Therefore, the selected method for producing the Persian question answering corpus was based on the collection strategy.

Table 1: Results of the cost-benefit method

Features	***Collection solution***	***Production solution***
Doing time	-3	-1
Research and executive costs	-3	-1
Required human resources	-3	-1
Technical and structural limitations	-1	-3
Verification capability	-1	-2
The size of the database	-3	-1
Domain comprehensiveness	-3	-1
Result	**-17**	**-10**

3-1-1- Identifying Reliable Websites

The next step after choosing solution 2 was to identify reliable websites that include a significant number of Persian questions and answers. A number of these websites were found by searching and the extracted sites were selected through two refinement stages and entered the final phase of corpus production. The first step was to refer to search engines to find websites having question and answer banks, and the second step was to look at the criteria for choosing the right website to crawl. The most important of these criteria are:

1) Intellectual property rights
2) Acceptable quantity
3) The possibility of crawling on the website
4) Random fact-checking of answers in domains
5) The comprehensiveness of the domain
6) Website ranking in Alexa[1]

The websites that entered the final phase, based on the above criteria, have been listed in Table 2.

Table 2: Persian websites suitable for crawling

Name	***#Q&As***	***Domain***	***Rank***
Hawzah	2371	Hawzah.net	460
Mamai	35609	Mamasite.ir	1119
Rasekhoon	83364	Rasekhoon.net	198
Shahab Moradi	7262	Shahab-moradi.ir	27747

3-2- Crawling Strategies

After selecting acceptable websites, the building process analysis was started with the aim of selecting the best way to extract questions and provide answers on each website. As observed in Fig. 1, the selection in this phase consists of four steps:

1) Manual extraction
   Considering the goal of producing a Persian question answering corpus with at least 50,000 records, this method was practically erosive and non-optimal and was removed from the selected strategies at the very beginning of the work.
2) Using ready-made tools
   Since the ready-made tools have limitations to crawl in all available websites, their use did not lead to the desired result. The main problem of these tools is data redundancy. Hence, this method was also rejected.
3) Production of a fully intelligent crawler robot
   At first glance, it seems like an attractive solution, however, the production of this robot may be a much more difficult project than the production of a Persian question answering corpus. Therefore, considering technical challenges, it is not possible to use this method.
4) Using an intermediate method (semi-intelligent)
   The only remaining option was this method which was selected and applied in this paper.

3-2-1- Semi-Intelligence Crawler (SIC)

As a new and unprecedented method, a crawler that is not a fully intelligent robot but can extract the materials required from the website using the human primary guide is designed. SIC is a revision crawler, whose primary setups for each website are easily carried out by a program, and performs the rest of the crawling steps, extracting and inserting into the database, in a completely intelligent way. Since the number of selected websites for producing Persian question answering systems was very small, performing the primary setting of the crawler was not a serious challenge, because it was important to configure it from any website.
To have a crawler that extracts exactly the desired information, it was necessary to examine each website separately with common methods, mostly innovative and sometimes combined methods. What is performed at this step is as follows:

1) View and check the page text
   Each page of the website that is loaded in the browser contains invisible information that can only be seen in the source text of that page. Information such as Document Type, Alphabetical coding, Metadata, Scripts used, Layout settings, File events, Local functions called, etc.
2) Parse Tree
   One of the most valuable ways to get the settings for each website is to rearrange the parse tree of different files on a website and discover the legal relationship between them.

Fig. 2 shows the metadata of the file. The useful part of designing SIC is the knowledge of the files that this page uses. These files are often used for either graphical settings of the page and have the suffix CSS (Cascading Style Sheets) or user-side functions that, if necessary,

[1] www.alexa.com/siteinfo

communicate with the server-side functions and are observed with the suffix JS (Java Script).

In Web 2.0 and later programming generations, an attempt is made to send fewer requests to the server, and with the minimum need to reload the entire page, the user requests are preprocessed on the user side by AJAX (Asynchronous JavaScript and XML) technology-based functions and then, sent to the server. By observing these functions and examining their structure, we were able to design a more powerful SIC website. In Fig. 3, these local functions are outlined.

In structured websites, fetching information from the order in their results' structure is a more efficient SIC design. Here, we sought to maximize the use of the order in the file structure. Fig. 4 shows a part of the parse tree.

3) Use of the web browser developer environments

The actions and reactions between the user system and the server on which the website is located are managed through browsers and are usually hidden from the ordinary user. To be aware of these interactions and behind-the-scenes events, you need to enter the web developer environment in the relevant browser.

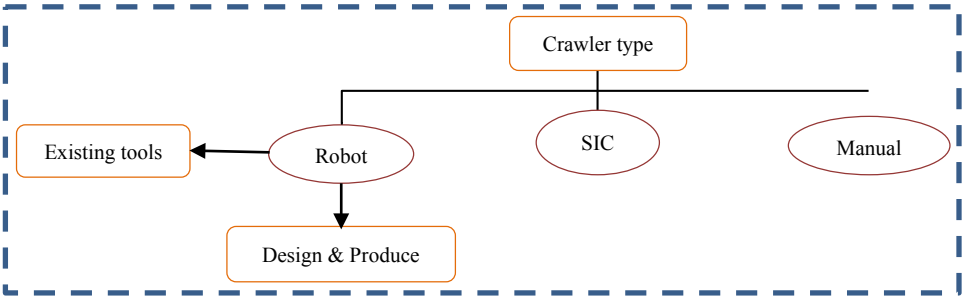


Fig. 1 Selection steps

```
<!DOCTYPE html>
<html>
<head><meta http-equiv="Content-Type" content="text/html; charset=utf-8">

  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="Author" content="Hadi Sharifian(09121869975)" />
  <meta name="Copyright" content="Copyright © 2018 Shimanaa, All Rights Reserved" />
  <meta name="robots" content="index, follow" />
  <meta name="keywords" content="شیمانا، آموزش، تکنولوژی، پیپاد سمپاش، موتلی کوپتر، هوش مصنوعی"/>
  <title>پیپاد سپرتگ</title>
  <!-- Tell the browser to be responsive to screen width -->
  <meta content="width=device-width, initial-scale=1, maximum-scale=1, maximum-scale=3.0, minimum-scale=0.86" name="viewport">
  <!-- Bootstrap 3.3.7 -->
  <link rel="stylesheet" href="dist/css/bootstrap-theme.css">
  <!-- Bootstrap rtl -->
  <link rel="stylesheet" href="dist/css/rtl.css">
  <link rel="stylesheet" href="dist/css/quiz.css">
  <!-- Font Awesome -->
  <link rel="stylesheet" href="bower_components/font-awesome/css/font-awesome.min.css">
  <!-- Ionicons -->
  <link rel="stylesheet" href="bower_components/Ionicons/css/ionicons.min.css">
  <!-- Theme style -->
  <link rel="stylesheet" href="dist/css/AdminLTE.css">
  <!-- AdminLTE Skins. Choose a skin from the css/skins
       folder instead of downloading all of them to reduce the load. -->
  <link rel="stylesheet" href="dist/css/skins/_all-skins.min.css">

  <!-- HTML5 Shim and Respond.js IE8 support of HTML5 elements and media queries -->
  <!-- WARNING: Respond.js doesn't work if you view the page via file:// -->
  <!--[if lt IE 9]>
  <script src="https://oss.maxcdn.com/html5shiv/3.7.3/html5shiv.min.js"></script>
  <script src="https://oss.maxcdn.com/respond/1.4.2/respond.min.js"></script>
  <![endif]-->
  <!-- jQuery 3 -->
  <script src="bower_components/jquery/dist/jquery.min.js"></script>
  <!-- Bootstrap 3.3.7 -->
  <script src="bower_components/bootstrap/dist/js/bootstrap.min.js"></script>
  <!-- SlimScroll -->
```

Fig. 2 Source text of a web page

```
<script>
    $('#addedtopics').hide();
    $('select').each(function(){
        ajaxlink="ajax/form_load_select.php?table="+$(this).attr("id");
        $(this).load(ajaxlink,function(){if($('#sections option').length>0&&$('#courses option').length>0&&$('#majors option').length>0&&$('#lessons
option').length>0)$('.modal-primary').hide();});
        $(this).css("width","95%");
    });
    $('select').bind('change',function(event){
        if($(this).attr("data-zir")!=""){
            $('.modal-primary').addClass('in');
            $('.modal-primary').css("display","block");
            tar=$(this).attr("data-zir");
            ajaxlink="ajax/form_load_zir.php?table="+$(this).attr("id")+"&zir="+$(this).attr("data-zir")+"&value="+$(this).val();
            $('#'+tar).load(ajaxlink,function(){$('.modal-primary').hide();});
            //$('.modal-primary').hide();
            //window.location='squiz#'+tar;
        }
    });
    $(document).on('change','#topics',function(){
        pm=$('#sections option:selected').text()+'-'+$('#courses option:selected').text()+'-'+$('#lessons option:selected').text()+'-'+$('#topics op
        pm+='<br><br>'+'این مبحث به آزمونک شما اضافه شود؟'+'<br><br>'+'<button type="button" class="btn btn-block btn-success">بله! اضافه شود</button>'+'<butto
block btn-danger">نه، منصرف شدم</button>';
        $('.modal-info .modal-body p').html(pm);
        $('.modal-info').addClass('in');
        $('.modal-info').css("display","block");

    });
    $(document).on('click','.btn-success',function(){
        txt='<li>'+$("#sections option:selected').text()+'-'+$('#courses option:selected').text()+'-'+$('#lessons option:selected').text()+'-'+$('#t
class="fa fa-fw fa-trash-o" id="'+$('#topics option:selected').val()+'"></i>'+'</li>';
        $('ol').append(txt);
        $('#topics option:selected').remove();
        $('.modal-info').hide();
        $('#addedtopics').show();

    });
    $('.modal-info').on('click','.btn-danger',function(){
        $('.modal-info').hide();
    });
    $('ol').on('click','.fa-trash-o',function(){
```

Fig. 3 Local functions called

```
<div class="attachment-block clearfix">
  <img class="attachment-img" src="images/arts/1589383615sirang-logo.jpg" alt="تماس با سیرنگ">
  <div class="attachment-pushed">
      <a href="9&تماس با سیرنگ" href="_blank"  title="تماس با سیرنگ">
      <h4 class="attachment-heading">تماس با سیرنگ</h4>
      </a>
      <div class="attachment-text">
          سیرنگ راه های ارتباطی متنوعی در اختیار شما قرار می دهد. فراخور ضرورت و نیاز خود می توانید یکی از آنها را انتخاب کنید.
  </div>
</div>
```

Fig. 4 A part of the parse tree

provides useful information to developers, such as:

- Details of each file
- The loading time of each element
- The protocol used to send the request to the server
- Return the data type of each request
- Variables exchanged between browser and server
- Displaying possible user side errors
- How to call the file updater function

Fig. 5 shows a part of the web developer environment in the Chrome browser (Google).

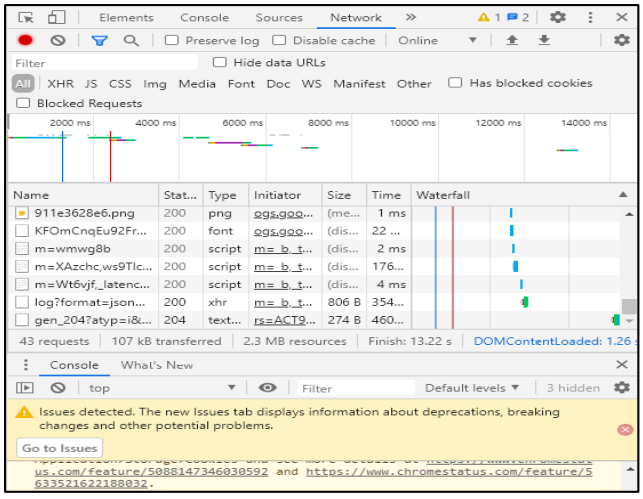


Fig. 5 A part of the web developer environment

4- SIC Strategy Implementation

In terms of implementation, to crawl the websites that we could extract the target of the project, the server-side programming language of PHP as well as the user-side programming languages of JS, jquery, and CSS were exploited. Plus, for inserting the extracted information into the crawls, MySQL was selected.

In the first crawl on the *.net* domain website, we faced an obstacle called Same-Origin Policy (SOP)[1]. SOP is a web security policy that does not allow web pages to load content from other domains, especially user-side scripts such as JS and jquery.

In fact, if a page in the *example1.com* domain tries to fetch content from the *example2.com* domain through one of the loads, post, get methods, or other common methods in various programming languages, browsers will be blocked according to agreed standards. The SIC crawler produced in the proposed approach was implemented in the hosting space of *shimanaa.com* and according to its mission, it was tasked to fetch and load the content of the question answering pages of the websites listed in Table 2 in this domain.

The next serious challenge which was tackled in this paper was facing this security policy.

Today, all websites have local scripts that are written primarily for the specific needs of the website. These scripts are stored in the web hosting space and addressing has been relatively defined in the context of their programs. For example, in the following code:

```
<script src=”bower_components/fastclick/lib/fastclick.js”></script>
```

As observed, the calling address of the *fastclick.js* file has been locally defined in the path of *bower_components/fastclick/lib/fastclick.js*.

Supposing there is one/more of these files on each website, they are necessary for the programs to run properly. However, SIC did not access any of these files on its host. Accordingly, there were two solutions to this problem:

1) Downloading all the required files of the websites and placing them in the same URLs.
2) Making changes to the fetched content so that the need for those files is eliminated.

Both solutions had some difficulties. The first one seemed to be a tedious task that required lots of time and operator work. The second one required an approach to pass the functions and procedures required by the websites.

Since our vision for the future was to carry out our approach on a larger scale, as well as to use the corpus produced in an operational plan, we selected the more difficult path and overcame this obstacle safely by inventing new methods.

According to W3C (World Wide Web Consortium) [2] standardization, every element in a web file must follow a set of rules. If this is not achieved, either the file upload process will be disrupted, or the functions and procedures will be called.

For example, the following code has a structural problem:

```
<li><a href=”profile_2_admin” target=_self ><i class=”fa fa-circle-o”></i>مشخصات کاربری.</li>
```

Because the tag of <a> had to be closed, which was not. The correct form of the above code is as follows:

```
<li><a href=”profile_2_admin” target=_self ><i class=”fa fa-circle-o”></i>مشخصات کاربری</a></li>
```

There were so many such cases that sometimes they seriously disrupted the SIC crawling process. To solve this problem, we made changes in the SIC to fetch as few elements as possible to make it easier to diagnose and fix structural defects.

Simultaneously with the successful crawling and fetching steps, the database entry step also should be performed. Hence, these three steps were implemented in two phases.

4-1- Phase 1: Crawling

In this step, a table was also needed to store information related to the question and answers rich pages. For this purpose, in the MySQL project database to store the information collected by SIC, 2 tables were created with the names *crl_links* and *crl_links_cats*. The *crl_links_cats* table stores classifications of questions and answers. The columns and some parts of the crawled records in this table are shown in Table 3.

Table 3: A part of the crl_links_cats table

cat_id	*subcat*	*Title*
1	2473	قرآن و تفسیر
2	2513	عقاید
3	2525	احکام
4	6330	مهدویت و انتظار
5	6699	تاریخ
6	2623	فرق، ادیان و مذاهب

The *crl_links* table stores the URLs of web pages containing questions and answers. The columns and some of the crawled records in this table are shown in Table 4. In Tables 3 and 4, *subcat* refers to the thematic classification code of each question and its reference website which has a unique code, shown in Table 4.

Table 4: A part of the crl_links table

link_id	*qid*	*subcat*	*Title*
1	1079865	2473	راه رهایی از دنیا دوستی چیست
2	723806	2473	مراد از فراز شریف الله مولی الذین آمنو…

[1] http://developer.mozilla.org/en-US/docs/Web/Security/Same-origin_policy
[2] https://www.w3.org/standards/

3	503363	2473	آیا نمونه‌هایی وجود دارد که نشان دهنده تاثیر فصاحت...
4	1079864	2473	چرا نماز باید به زبان عربی خوانده شود
9	1079847	2473	اگر اعمال دنیوی ما مربوط به تصمیمان در عالم ذر....
24	1079733	2473	معنای اسلام چیست
25	1079732	2473	در زندگی‌ام همیشه با مشکلات و گرفتاری روبه‌رو هستم...

4-2- Phase 2: Insert Fetch

At this point, the SIC crawler, based on the information in the *crl_links* table, scanned the page of each address in this table and entered the information for each question and answer in the *crl_questions* table. The columns and some of the crawled records in this table are shown in Table 5.

5- Evaluation

After the production of the corpus, the only thing left to do was to standardize it as shown in Fig. 6.
The indicators considered are:

- The text of the questions and answers should be free of any HTML protocol tags and symbols.
- Having no nature other than questions and answers.
- The data should not be duplicated.
- The addresses must be valid for fact-checking.
- Classifications should not be too general or partial.

The pre-standardization corpus contained 83, 364 questions and answers. At this step, duplicate questions were removed first. Then, in a separate table, we saved a copy of the corpus after deleting the stop word. We received the list of stop words from this link.
After processing all the questions and answers that were in the draft corpus, we proceeded to extract all the unique words. 56,925 of these words were stored in a separate table, in which, 4 parameters were calculated for each word and the table was updated. In this regard, we used the four metrics from Eq. (1) to Eq. (4), [25].

$$\mathrm{TF(t,d)} = \frac{\text{Number of times term t appears in a document}}{\text{Total number of terms in the document}} \quad (1)$$

$$\mathrm{DF(t)} = \frac{\text{Number of documents that the term appears in them}}{\text{Total number of documents}} \quad (2)$$

$$\mathrm{IDF(i)} = \frac{\ln(N+1)}{\mathrm{DF(i)}} + 1 \quad (3)$$

Where N is the total number of documents in the collection.

$$\text{TF-IDF(i,d)} = \text{TF(i,d)} \times \text{IDF(i)} \quad (4)$$

The corpus includes rather huge amounts of information in question-and-answer form; therefore, suitable metrics for evaluation, are those used in information retrieval. There are several of them, but order-aware metrics are chosen in this paper as follows (Eq. (5) to Eq. (10)) [10, 26, 27, 11]:

a. Mean Reciprocal Rank (MRR)

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{r_i} \quad (5)$$

Where |Q| denotes the total number of queries and r_i shows the rank of the *i*th relevant result.

b. Average Precision (AP)

$$AP = \frac{\sum_{i=1}^{n}\left(P(i)\times rel(i)\right)}{number\ of\ relevant\ items} \quad (6)$$

Where *P(i)* is the Precision@k metric and *rel(i)* is 1 if the *i*th item is relevant otherwise is 0.

c. Mean Average Precision(MAP)

$$MAP = \frac{1}{|Q|}\sum_{i=1}^{|Q|}AP(i) \quad (7)$$

Where |Q| denotes the total number of queries and *AP(i)* is the average precision for the *i*th query.

d. Cumulative Gain (CG@k)

$$CG@k = \sum_{i=1}^{k} rel_i \quad (8)$$

e. Discounted Cumulative Gain (DCG@k)

$$DCG@k = \sum_{i=1}^{k}\frac{rel_i}{\log_2(i+1)} \quad (9)$$

f. Normalized Discounted Cumulative Gain (NDCG@k)

$$NDCG@k = \frac{DCG@k}{IDCG@k} \quad (10)$$

Where *IDCG@k* is the ideal *DCG@k* (more relevant item comes first).
Metrics *CG@k*, *DCG@k*, and *NDCG@k* consider the grade of relevancy while the first three metrics do not mention it.

Table 5: Part of the crl_questions table

id	*question*	*answer*	*keyword*
1	راه رهایی از دنیا دوستی چیست	اجمالی: دنیا مونث....	رهایی از دنیادوستی | دنیادوستی | قرآن مجید | زندگی آخر ...
2	مراد از فراز شریف الله مولی الذین آمنو...	آیه شریف "ذلک بان....	مولا | مومنان | کافران
3	آیا نمونه‌هایی وجود دارد که نشان دهنده تاثیر فصاحت...	آری، نمونه‌های فراوانی در این...	
4	چرا نماز باید به زبان عربی خوانده شود	برای روشن شدن پاسخ، ابتدا...	نماز | حقوق و احکام | زبان عربی | نماز با زبان عربی
9	اگر اعمال دنیوی ما مربوط به تصمیمان در عالم ذر....	شرح در مورد مسئله...	عالم ذر | دنیا | اعمال دنیوی | تفسیر | جبر یا اختیار | عدالت...
24	معنای اسلام چیست	پاسخ اسلام در لغت...	معنای اسلام | اسلام | دین اسلام | قرآن کریم | توحید کامل
25	در زندگی‌ام همیشه با مشکلات و گرفتاری روبه‌رو هستم...	شرح در کاری که از....	ضرر و بیماری | امتحان | عدالت| پروردگار | ابتلا و امتحان...

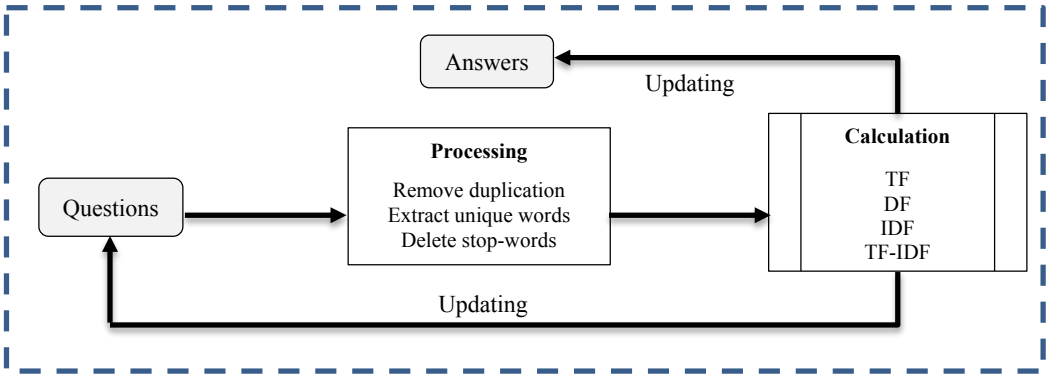


Fig. 6 Schematic of corpus standardization steps

To evaluate the corpus, we ran a list of questions written by a user interface in PHP in the MySQL database. Some of these questions include:

1) ازدواج با اهل کتاب چه حکمی دارد؟
2) چرا هنگام مسواک زدن لثه ام خونی می‌شود؟
3) دختر از پدر چه مقدار ارث می‌برد؟
4) نماز خواندن با لباس خونی چه حکمی دارد؟
5) چگونه درد زایمان را تحمل کنیم؟

These queries were edited before being entered as queries and their stop words were removed. The output of questions 3 and 4 are shown in Tables 6 and 7.

Table 6: Results of the evaluations made on question 4

نماز خواندن با لباس خونی چه حکمی دارد؟					
حکمی	خونی	لباس	خواندن	نماز	
2031	128	768	807	4213	TF
1969	109	616	747	3009	DF
3,42461	6,31854	4,58664	4,39382	3,0005	IDF
3,53244	7,41993	5,71841	4,74674	4,20113	TF-IDF

Table 7: Results of the evaluations made on question 3

دختر از پدر چه مقدار ارث می برد؟				
می برد	ارث	پدر	دختر	
295	515	1368	1459	TF
279	438	1073	1017	DF
5,37867	4,92767	4,03167	4,08527	IDF
5,68713	5,79395	5,1401	5,86078	TF-IDF

The evaluation of the two mentioned questions had acceptable results, according to the predefined metrics.

In brief, in the proposed corpus, questions that are entered directly from the questions in the system as input by the user interface are retrieved with 100% accuracy. Questions that use a part of the vocabulary contained in the system questions lead to 100% accurate retrieval. Questions that are randomly generated by a human agent have a wide range of retrieval accuracy from bad to very good, depending on the type of vocabulary used and the quantity of the corresponding 4 parameters. More precisely, if the user input keywords are available in the database, the results will be very good, otherwise, the system output may not be good.

6- Experimental Results

As described in the previous section, after eliminating the stop words from 83,364 questions, 56,925 unique words remained. The first step of evaluating the corpus is designing a robust model, which is useful for the whole corpus rather than part of it. Therefore, the stored words in the MySQL database were analyzed with a focus on Tf-Idf and Df parameters. Table 8 displays the range of changes in these parameters.

Table 8: Range of changes in Tf-Idf and Df

Item	*Minimum value*	*Maximum value*
Tf-Idf	2.9	66.1
Df	1	3522

The scattered values of these parameters led to the design of two evaluation models that will be discussed in the following.

6-1- Model No.1

Designing, implementing, and evaluating this model have been done within 10 steps:

1. Dividing unique words into five clusters according to their Tf-Idf.
2. Calculating the average Tf-Idf for each cluster.
3. Selecting the sample word from the database randomly based on the average Tf-Idf in each cluster.
4. Creating an arbitrary question by a human agent using the selected word.
5. The created question is run as a query in the database by means of the designed UI
6. The results are fetched based on the model relevancy prediction.
7. Duplicate results are removed, and the first 5 results are retained. (k=5)
8. Based on a ground-truth annotation each result is assigned a score between 1 (least relevant) and 5 (most relevant).
9. The metrics are calculated for each query.
10. The quality of evaluation is determined by examining the calculated metrics.

Table 9 shows the unique words division in 5 clusters and their sample words.

In Table 10, cells with green color have a grade upper than 2 and are considered as a True result. Besides, cells with red color have a grade lower than 3 and are considered as a False result. Then, in Table 11, the standard metrics for Model No.1 are calculated.

6-2- Model No.2

This model is like Model No.1, just it uses df parameters instead of Tf-Idf.

1. Dividing unique words into five clusters according to their df values.
2. Calculating the average df for each cluster.
3. Selecting the sample word from the database randomly based on the average df in each cluster.
4. Creating an arbitrary question by a human agent using the selected word.
5. The created question is run as a query in the database by means of the designed UI.
6. The results are fetched based on the model relevancy prediction.
7. Duplicate results are removed, and the first 5 results are retained (k=5).
8. Based on a ground-truth annotation each result is assigned a score between 1 and 5 (most relevance).
9. The metrics are calculated for each query.
10. The quality of evaluation is determined by examining the calculated metrics.

Table 12 shows the unique words division in 10 clusters and their sample words. In Table 13, cells in green have a grade upper than 2 and are considered as a True result. Besides, cells in red have a grade lower than 3 and are considered a False result. Then, in Table 13, the standard metrics for Model No.1 are calculated.

Table 9: Dividing the unique words into 5 clusters and choosing sample words

cluster	*Tf-Idf range*	*Tf-Idf average*	*Selected Word*	*Pronunciation*	*Meaning*	*Tf-Idf*
1	2.9 – 8.3	7.239	یقین	/yagheen/	Certainty	6.1
2	8.4 – 13.7	10.53	ازدواج	/ezdevaaj/	Marriage	12.3
3	13.8 – 19.2	15.524	طلسم	/telesm/	talisman	18.8
4	19.3 – 24.9	21.81	تبعید	/tab'eed/	Exile	21.9
5	25 – 66.1	35.98	روح	/ruh/	ghost	25.9

Table 10: Ground-Truth Annotation given grade to the results

Selected Word	*Arbitrary Question*	*Grades 1 (least relevant) to 5 (most relevant)*				
		Result1	Result2	Result3	Result4	Result5
یقین	چگونه به یقیین برسیم؟	5	2	1	1	1
ازدواج	ازدواج با اهل کتاب چه حکمی دارد؟	5	4	5	3	4
طلسم	چگونه می توان طلسم را باطل کرد؟	5	3	2	4	5
تبعید	در چه صورت فرد تبعید می شود؟	4	4	4	5	5
روح	مشکلات روحی چگونه درمان می شود؟	4	1	5	2	4

Table 11: Calculated Metrics for Model No.1

Selected Word	*Reciprocal Rank*	*Average Precision*	*CG@5*	*DCG@5*	*NDCG@5*
یقین	1	1	10	7.581	1
ازدواج	1	1	21	12.867	0.987
طلسم	1	0.888	19	11.555	0.945
تبعید	1	1	22	12.617	0.940
روح	1	0.756	16	9.543	0.886

Table 12: Dividing the unique words into 10 clusters and choosing sample words

Cluster	*Df range*	*Df average*	*Selected Word*	*Pronunciation*	*English Equivalent*
1	11 - 20	15	خودسازی	/khodsaazi/	Self-construction
2	21 - 30	25	اهانت	/ehaanat/	contempt
3	31 - 40	35	مسئولیت	/mas'uliat/	responsibility
4	41 - 50	45	جوراب	/juraab/	socks
5	51 - 60	55	خرما	/khorma/	Date palm
6	61 - 70	65	پوسیدگی	/puseedegi/	decay
7	71 – 80	76	لاغر	/laaghar/	thin
8	81 – 90	86	امانت	/amaanat/	trusteeship
9	91 – 100	95	صدقه	/sadagheh/	alms
10	100 - 3522	332	زکات	/zakat/	zakat

Table 13: Ground-Truth Annotation given grade to the results

Selected Word	*Arbitrary Question*	*Grades 1(least relevant) to 5(most relevant)*				
		Result1	Result2	Result3	Result4	Result5
خودسازی	خودسازی چگونه انجام می شود؟	3	5	5	4	5
اهانت	اهانت به اهل سنت چه حکمی دارد؟	1	5	5	2	1
مسئولیت	مسئولیت والدین درباره فرزندان چیست؟	5	4	5	5	1
جوراب	نظر اسلام درباره جوراب پوشیدن خانم ها چیست؟	5	2	5	5	5
خرما	خرما خوردن چه فایده ای دارد؟	5	5	3	4	4
پوسیدگی	نشانه پوسیدگی دندان چیست؟	5	5	4	5	5
لاغر	چطور لاغر شوم؟	1	5	2	5	4
امانت	احکام مربوط به خیانت در امانت چیست؟	5	4	5	5	5
صدقه	چه کسی مستحق دریافت صدقه است؟	5	5	3	2	2
زکات	به چه چیزهایی زکات تعلق می گیرد؟	3	5	5	5	5

Table 14: Calculated Metrics for Model No.2

Selected Word	*Reciprocal Rank*	*Average Precision*	*CG@5*	*DCG@5*	*NDCG@5*
خودسازی	1	1	22	12.317	0.910
اهانت	0.5	0.583	14	7.904	0.793
مسئولیت	1	1	20	12.567	0.984
جوراب	1	0.804	22	12.855	0.946
خرما	1	1	21	12.929	0.991
پوسیدگی	1	1	24	14.248	0.992
لاغر	0.5	0.533	17	8.860	0.777
امانت	1	1	24	14.117	0.983
صدقه	1	1	17	11.292	1.000
زکات	1	1	23	12.748	0.912

In both models, while calculating the Reciprocal Rank (RR) and Average Precision (AP), results with grades lower than 3, are considered as a False result and True otherwise. Table 15 represents the comparison of the evaluation results of the two models.

Table 15: Comparing evaluation results of two models

Model	***MRR***	***MAP***	***NDCG@5***
No.1	1	0.929	0.951
No.2	0.9	0.892	0.929

Comparison of the evaluation results of the 2 models is as expected, as the parameter used in Model No.1 is more effective than the parameter used in Model No.2. Figs. 7 and 8 reveal more details from clusters in Model No.1.

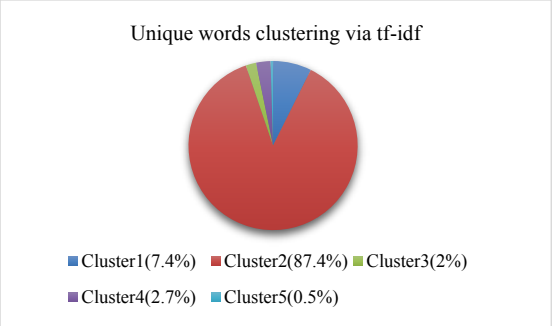


Fig. 7 Unique words clustering via Tf-Idf

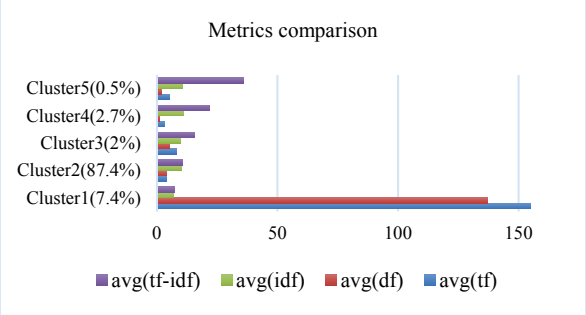


Fig. 8 Metrics comparison between five clusters of Model No.1

7- User Interface

Now, this corpus with 53,844 Persian questions and answers and the name "Popfa", has been provided to the interested parties and researchers. A part of the User Interface (UI) designed for it is shown in Fig. 9. In the user interface, questions that are entered as queries are returned based on two types of searches:

1) (Search by questions) جستجو بر اساس پرسش‌ها
2) (Search by answers) جستجو بر اساس پاسخ‌ها

The output of the UI is shown in Fig. 9, also this interface is available to the public through this link. For each question, the output is presented to the user as two separate lists based on the similarity to the questions and answers available in the created corpus.

Fig. 9 The designed UI

8- Discussion

Considering the number of questions in the corpus and the various classifications of it, it has a significant advantage over the few existing systems in the Persian language, both in terms of quantity and quality. It has 2 general classifications, Religious and Medical. Each class is divided into other sub-classes. Table 16 shows some of the thematic classifications of the questions and answers of the corpus. 20 different classifications in the designed system are presented in Table 16.

Table 16: Different classifications in the corpus

cat_id	***title***	***cat_id***	***title***
1	قرآن و تفسیر	11	اندیشه اسلامی
2	عقائد	12	حدیث شناسی
3	احکام	13	منطق
4	مهدویت و انتظار	14	فلسفه
5	تاریخ	15	کلام
6	فرق، ادیان و مذاهب	16	عرفان و تصوف
7	تربیت و مشاوره	17	حقوق
8	دین پژوهی	18	مسائل زنان
9	اخلاق	19	پزشکی
10	سیاست	20	علمی

The importance of addressing different simple and challenging question types and scenarios in the field of question answering systems is undeniable. It should be noted that Popfa corpus contains various kinds of questions including descriptive, factoid, confirmation, comparative, relationship-based, and list questions. However, the generated corpus does not include

hypothetical and complex questions. Hypothetical questions ask for information associated with any hypothetical event and no specific answers to these questions are necessarily available. For example:
What will happen if a big earthquake occurs in Tehran?
Complex questions are more challenging to answer, and their answers generally consist of a list of different kinds of answers. For instance:
What are the reasons for heavy traffic in megacities?
These questions can have various answers according to the idea of each person.
Here, to ensure ease of access for readers, we have made the corpus available on GitHub, where it can be downloaded and utilized for research and development purposes. The link to access the Popfa Persian Question Answering corpus can be found in this link.

9- Conclusion and Future Works

In this paper, some weaknesses and challenges of the Persian language question answering systems have been highlighted. Then, Popfa question answering corpus has been generated due to a standard procedure to fulfill the mentioned need for Persian NLP tasks. In this regard, different methods for generating a question answering corpus are investigated. Eventually, the innovative method of designing a semi-intelligent crawler in this paper has led to the production of a corpus containing 53,844 questions in Persian. Furthermore, by performing the corpus standardization processes and designing a user interface for better communication between the audience and the system, we achieved a set of standard and reliable questions and answers. In the last stage, we compared the Popfa corpus with the corpora used in the related question answering systems in the Persian language, in which many questions of the Popfa corpus and its sub-thematic diversity were evaluated as a competitive advantage.
Undoubtedly, the development and improvement of Persian question answering systems to handle all ranges of questions would require further research and implementation efforts. As the Popfa corpus contains various kinds of questions, future research can work on more complicated types of questions, such as hypothetical and complex questions. In this regard, exploring and explaining the potential approaches, techniques, and models that could be employed to tackle more complicated questions would indeed be a valuable area for future research. Additionally, the domain of questions can be diversified and expanded. Moreover, the user interface designed for Popfa can be developed into an intelligent robot. In the future, this user interface can be turned into a powerful text assistant using various machine learning methods. On top of that, the link to the created corpus is provided, so researchers can apply it in Persian question answering systems in the future.

Acknowledgments

The authors acknowledge that this study is edited by Dr. Belmond Yoberd, Divisional Director, Mott MacDonald Limited, United Kingdom.

References

[1] R. French, "The Turing Test: The first 50 years," Trends in Cognitive Sciences, vol. 4, no. 3, pp. 115-122, 2000.
[2] Z. Khalifeh Zadeh, and M. A. Zare Chahooki, "An Effective Method of Feature Selection in Persian Text for Improving the Accuracy of Detecting Request in Persian Messages on Telegram," Journal of Information Systems and Telecommunication (JIST), vol. 8, no. 32, pp. 249-262, 2021.
[3] N. Tohidi, and S. M. H. Hasheminejad, "A Practice of Human-Machine Collaboration for Persian Text Summarization", in The 27th International Computer Conference, Tehran, 2022.
[4] A. Hoseinmardy, and S. Momtazi, "Recognizing Transliterated English Words in Persian Texts", Journal of Information Systems and Telecommunication (JIST), vol. 8, no. 30, pp. 84-92, 2020.
[5] N. Tohidi, C. Dadkhah, and R. B. Rustamov, "Optimizing Persian multi-objective question answering system," International Journal on Technical and Physical Problems of Engineering (IJTPE), vol. 13, no. 46, 2021.
[6] M. Breja, "A Customized Web Spider for Why-QA Pairs Corpus Preparation," Journal of Information Systems and Telecommunication (JIST), vol. 11, no. 41, pp. 41-47, 2023.
[7] Tohidi, Nasim., Dadkhah; Chitra. Rustamov, and Rustam B., "Optimizing the Performance of Persian Multi-objective question answering system", in The 16th International Conference on Technical and Physical Problems of Engineering, Istanbul, Turkey, 2020.
[8] C. P. Masica, The Indo-Aryan Languages, New York, Cambridge University Press, 1993.
[9] D. Khashabi, A. Cohan, S. Shakeri, P. Hosseini, P. Pezeshkpour, M. Alikhani, M. Aminnaseri, M. Bitaab, F. Brahman,S. Ghazarian, M. Gheini, A. Kabiri, R. Karimi Mahabagdi, O. Memarrast, et al., "ParsiNLU: A Suite of Language Understanding Challenges for Persian," Transactions of the Association for Computational Linguistics, vol. 9, p. 1147–1162, 2021.
[10] E. M. Voorhees, "The TREC-8 Question Answering Track Report (1999)," in In Proceedings of TREC-8, 1999.
[11] N. Tohidi, and S. M. H. Hasheminejad, "MOQAS: Multi-objective question answering system", Journal of Intelligent & Fuzzy Systems, vol. 36, no. 4, pp. 3495-3512, 2019.
[12] I. Khodadi, and M. Saniee Abadeh, "Genetic programming-based feature learning for question answering," Elsevier, Information Processing and Management, vol. 40, 2015.
[13] M. Joshi, E. Choi, D. Weld, L. Zettlemoyer, "TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017.

[14] A. Muttaleb Hasan, and L. Q. Zakaria, "Question classification using support vector machine and pattern matching," Journal of Theoretical and Applied Information Technology , vol. 87, no. 2, pp. 259-265, 2005.
[15] H. Veisi, and H. Fakour Shandi, "A Persian Medical Question Answering System," International Journal on Artificial Intelligence Tools, vol. 29, no. 6, 2020.
[16] A. Aleahmad, H. Amiri, E. Darrudi, and F. Oroumchian, "Hamshahri: A standard Persian text collection", Knowledge-Based Systems, vol. 22, no. 5, pp. 382-387, 2009.
[17] A. Mollaei, S. Rahati Quchani, and A. Estaji, "Question classification in Persian language based on conditional random fields", in 2nd International eConference on Computer and Knowledge Engineering (ICCKE), 2012.
[18] E. Sherkat, and M. Farhoodi, "A Hybrid Approach for Question Classification in Persian Automatic Question Answering Systems", in 4th International eConference on Computer and Knowledge Engineering (ICCKE), Mashahd, Iran, 2014.
[19] A. P. Ben Veyseh, "Cross-Lingual Question Answering Using Common Semantic Space", in Proceedings of the 2016 Workshop on Graph-based Methods for Natural Language Processing, San Diego, California, 2016.
[20] Y. Boreshban, H. Yousefinasa, and S.A. Mirroshandel, "Providing a Religious Corpus of Question Answering System in Persian", Signal and Data Processing, vol. 15, no. 1, pp. 87-102, 2018.
[21] R. Etezadi, and M. Shamsfard, "PeCoQ: A Dataset for Persian Complex Question Answering over Knowledge Graph", in 11th International Conference on Information and Knowledge Technology (IKT), Tehran, Iran, 2020.
[22] N. Abadani, J. Mozafari, A. Fatemi, M. A. Nematbakhsh, and A. Kazemi, "ParSQuAD: Persian Question Answering Dataset based on Machine Translation of SQuAD 2.0", International Journal of Web Research, vol. 4, no. 1, pp. 34-46, 2021.
[23] A. Kazemi, J. Mozafari, and M. A. Nematbakhsh, "PersianQuAD: The Native Question Answering Dataset for the Persian Language", IEEE Access, vol. 10, pp. 26045-26057, 2022.
[24] K. Darvishi, N. Shahbodagh, Z. Abbasiantaeb, and S. Momtazi, "PQuAD: A Persian Question Answering Dataset", arXiv:2202.06219, 2022.
[25] D. Jurafsky, and H. Martin James, Speech and Language Processing, Upper Saddle River, NJUnited States: Prentice Hall, 2019.
[26] R. Dragomir; H. Qi, H. Wu, and W. Fan, "Evaluating Web-based Question Answering Systems", in The Third International Conference on Language Resources and Evaluation (LREC'02), Las Palmas, Canary Islands - Spain, 2002.
[27] K. Järvelin, and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," ACM Transactions on Information Systems, vol. 20, no. 4, pp. 422-446, 2002.

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 152-161
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Whispered Speech Emotion Recognition with Gender Detection using BiLSTM and DCNN

Aniruddha Mohanty[1*], Ravindranath C Cherukuri[1]

[1].Department of Computer science and Engineering, Christ (Deemed to be University), Bangalore, India

Received: 22 Aug 2023/ Revised: 04 Apr 2024/ Accepted: 21 May 2024

Abstract

Emotions are human mental states at a particular instance in time concerning one's circumstances, mood, and relationships with others. Identifying emotions from the whispered speech is complicated as the conversation might be confidential. The representation of the speech relies on the magnitude of its information. Whispered speech is intelligible, a low-intensity signal, and varies from normal speech. Emotion identification is quite tricky from whispered speech. Both prosodic and spectral speech features help to identify emotions. The emotion identification in a whispered speech happens using prosodic speech features such as zero-crossing rate (ZCR), pitch, and spectral features that include spectral centroid, chroma STFT, Mel scale spectrogram, Mel-frequency cepstral coefficient (MFCC), Shifted Delta Cepstrum (SDC), and Spectral Flux. There are two parts to the proposed implementation. Bidirectional Long Short-Term Memory (BiLSTM) helps to identify the gender from the speech sample in the first step with SDC and pitch. The Deep Convolutional Neural Network (DCNN) model helps to identify the emotions in the second step. This implementation is evaluated using the wTIMIT data corpus and gives 98.54% accuracy. Emotions have a dynamic effect on genders, so this implementation performs better than traditional approaches. This approach helps to design online learning management systems, different applications for mobile devices, checking cyber-criminal activities, emotion detection for older people, automatic speaker identification and authentication, forensics, and surveillance.

Keywords: Whispered Speech; Emotion Recognition; Speech Features; Data Corpus; BiLSTM; DCNN.

1- Introduction

Emotions are the humans' short-lived feelings that affects thinking, actions, relationships, and social interactions. Emotions express humans' physiological and emotional states with facial expressions and body language. Whispered speech is a form of speech that expresses emotions. Whispered speech is a type of communication produced with breath without any noise and excitation of voice. The whispered speech structure changes significantly because of the lack of periodic excitation in the voice. This results in missing speech and reduced transparency in communication.

The difference between normal and whispered speech is the absence of vocal tract vibrations due to the vocal tract's physiological blocking. The strength of whispered speech is minute and without voice compared to phoned speech. The spectral and prosodic features help to detect the whispered speech. Prosodic features of speech vary over time. Spectral features of whispered speech are highly accurate over unvoiced consonants, voiced consonants, and formants in vowels [1].

The impacts on Speech Emotion Recognition (SER) are due to various acoustic conditions such as compressed, noisy, telephonic conversations and imitator speeches. Other environmental conditions, such as stress, rhythm, and intonation, can affect SER. Whispered speech is also affected by similar acoustic and environmental conditions. SER depends on acoustic characteristics and gender in some scenarios. Hence gender plays a vital role in SER.

Nowadays, several whispered speech samples of male and female voices are available. The effectiveness of the SER is more when the implementation is in two parts. The identification of the gender [2] happens initially using Bidirectional Long Short-term Memory (BiLSTM) with speech features. The next step is detecting emotions using various speech features [3] and Deep Convolutional Neural Network (DCNN).

The structure of the paper includes various sections. Section 2 describes Human Emotions and their Applications. Details of the Related Work are in Section 3. Section 4 describes the System Model, which is the black box view of the implementation. Section 5 is the Model

Design that describes the details of the speech features and the deep learning models used in this implementation. Section 6 gives the Experiment and Result Analysis. Section 7 briefly discuss the Conclusion and Future Work.

2- Human Emotions and their Applications

Human emotions are the mental state caused by countless associated views, feelings, behavioral replies, and the degree of pressure and annoyance. It is often associated with attitude, temperament, behavior, disposition, and creativity [4]. Emotion recognition helps to detect a humans' emotional mood, which lasts hours and days. Speech conveys emotions such as anger, disgust, fear, happiness, neutrality, sadness, and surprise. The machine automatically detects various emotions from speech with the help of different algorithms. Emotion Recognition helps to:

- Detect customers' intentions based on the teleconference.
- Detect cybercrime.
- Students' attention and teachers' content adjustment.
- Disability assistance
- Customer satisfaction
- Stress monitoring
- Social media analysis
- Suspicious activity
- Human-machine interaction and so on.

3- Related Work

Over time, numerous studies have detected emotions from whispered speech. The various deep learning models detect emotions based on the speech features extracted from the collected whispered speech data corpus. The SER for normal and whispered speech is diverse because of vocal excitation. Emotions vary with many factors; gender is among the most influential factors [5]. Identifying gender in the first step improves emotion detection from whispered speech. So, the related work explored is on gender detection and emotion recognition from whispered speech.

MFCC obtained by the Hilbert envelope approach and weighted instantaneous frequencies (WIFs) obtained by the coherent demodulation help to detect gender in whispered speech samples [6]. There is an opportunity to explore gender detection in noisy speech conditions using these approaches.

Autoencoder-enabled features in the transfer learning framework propose to practice phonated data to identify emotions from Whispered speech [7]. The feature extraction is from the Geneva Whispered Emotion Corpus (GeWEC) and Berlin Emotional Speech Database (EMO-DB) data corpus. The acoustic features such as Mel-frequency cepstral coefficient (MFCC), root mean square (RMS), frame energy, zero-crossing-rate (ZCR), pitch frequency (F0), probability of voicing autocorrelation function are the inputs to evaluate the framework. Implementing deep learning concepts on spontaneous data gives more accuracy than the current framework.

Gender is detected to target an anonymous speaker. The Deep Neural Network (DNN) model is used to generalize gender by using the MFCC speech feature. This implementation is applied to the wTIMIT dataset to verify the gender and recognize the speaker [8]. The DNN model cannot generalize gender; evaluation can happen with other datasets.

MFCC and CNN, with the fully connected network, detect gender and emotions like anger, disgust, fear, happiness, sadness, surprise, and a neutral state [9]. The concept verification happens on RAVDESS, CREMA-D, SAVEE, and TESS datasets, having an accuracy of 92.283%, which is better than the traditional model.

The final prediction of the implemented model is to learn the mutually spatial-spectral features happens by a two-stream deep convolutional neural network with an iterative neighborhood component analysis (INCA) and the most discriminatory optimal features [10]. The concept verification happens on EMO-DB, SAVEE, and RAVDESS emotional speech corpora which perform with 95%, 82%, and 85% accuracy rates. Real-time applications with natural and huge data corpora can help to extend this concept to identify emotions.

Mel Frequency Magnitude Coefficient (MFMC) and three spectral features, namely MFCC, log frequency power coefficient, and linear prediction cepstral coefficient, are used with the help of Support Vector Machine (SVM) modeling [11]. The performance evaluation uses this concept on Berlin, RAVDESS, SAVEE, EMOVO, and eNTERFACE data corpus. Feature selection, feature fusion, and multiple classification schemes improve the performance of MFMC.

The proposed MFF-SAug research in which noise removal improves emotion. White Noise Injection, Pitch Tuning techniques, MFCC, ZCR, RMS speech features, and Convolutional Neural Network (CNN) modeling [12] detect emotions. Emotion detection during the interaction between people can extend the MFF-SAug approach.

The proposed Neural Network-based Blended Ensemble Learning (NNBEL) model is composed of a 1-dimensional Convolution Neural Network (1D-CNN), Long Short-Term Memory (LSTM), and CapsuleNets. LSTM receives input from the Log Mel-spectrogram speech features, while 1D-CNN and CapsuleNets receive input from the MFCC. Each model's output is fed to Multi-Layer Perceptron (MLP) and predicts the final emotions [13]. This model shows 95.3% and 94% accuracy on RAVDESS and IEMOCAP datasets, respectively.

TrustSER [14] implemented a general framework to determine the SER system's trustworthiness using deep learning techniques. Trustworthiness evaluates privacy (gender information, speaker demographics), safety, fairness, sustainability, and emotions (sad, angry, happy, neutral). The architecture of the TrustSET framework uses CNN encoder and Transformer encoder models. Trustworthy profiles under the Federated Learning scenario might improve privacy, fairness, and safety.

A hybrid meta-heuristic ensemble-based classification [15] helps to detect speech emotions. Raw speech samples are filtered using the Butterworth filter; then, spectrogram speech features are extracted from each frame to create a hybrid feature vector. The ensemble-based classification is applied to hybrid feature vectors to classify emotions. The ensemble-based classification contains a Recurrent Neural Network (RNN), a Deep Belief Network (DBN), and an Artificial Neural Network (ANN).

The above-related work shows that it is crucial to identify gender to segregate emotions in a speech. Male and female emotions are different based on situations. Whispered speech is an essential concept of emotion recognition. There is an opportunity to experiment further based on other datasets of whispered speech. So, this is a motivation to work on emotion recognition in whispered speech, as this is a less explored area of research.

4- System Model

The analysis of emotion recognition from the whispered speech segregates into two parts. One is gender detection, and the second is emotion detection as shown in Fig 1.

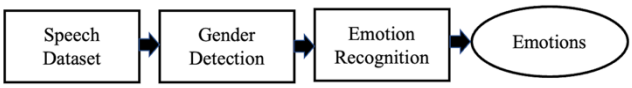


Fig. 1 Emotion recognition from whispered speech

As part of gender detection, pre-processing of the speech samples is performed, and then pitch and SDC are extracted as part of prosodic and spectral features, respectively. Both the features are fused to get a single feature set with the help of multifeature fusion. PCA helps to reduce the dimensions of extracted features and is used as input to Bidirectional Long short-term memory (BiLSTM) to classify genders, as shown in Fig 2.

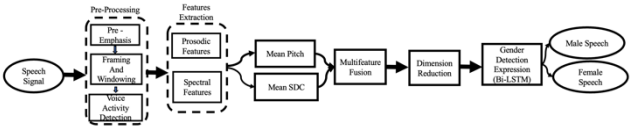


Fig .2 Gender detection from whispered speech

Following Gender detection, the augmentation of speech data helps to get more realistic data. Then the speech features such as Chromatogram, Zero-Crossing Rate, Spectral Central, MFCC, Spectral Flux, and Mel Spectrograms are extracted. All the extracted features combine to get a single feature using multifeature fusion. Then the dimensions are reduced to get the optimal data points. The data points are inserted into the Deep Convolutional Neural Network (DCNN) to identify the emotions, as shown in Fig 3.

5- System Design

The implementation of system design happens in two parts. The first part identifies the male and female gender from the whispered speech samples. Then the determination of different emotions like sadness, happiness, fear, anger, surprise, disgust, and neutral are detected in gender-segregated speech samples.

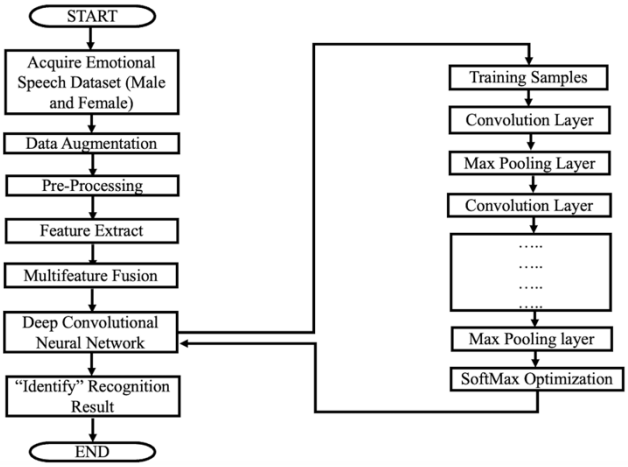


Fig. 3 Emotion Recognition from Male and Female whispered speech.

5-1- Whispered Speech Data Corpus

wTIMIT [16] is a whispered voice data corpus having 450 phonetically balanced sentences with 29 speakers. There are 11,324 utterances with a normal voice, which can be used for normal and whispered training. This data is available in two parts - train and test divisions. The samples contain an 8 kHz sampling rate with a high-quality and low-pass filter.

5-2- Data Augmentation

Data augmentation [17] is the method of getting more realistic data from the existing data, which helps to add more training data to the model, reduce overfitting, and increase the model's generalization ability. Below are the steps to generate synthetic data.

- **Shifting Time:** Shifting time is a simple concept where the audio shifts to the left or right with a haphazard second. If the audio shifts to the left

with z seconds, then the first z seconds are marked as silence. Similarly, if the audio shifts to the right with z seconds, the last z seconds are marked as silence.

- **Changing Pitch:** Pitch change is adjusting the pitch randomly without upsetting the speediness of the audio file.
- **Changing Speed:** Speed audio changes by stretching the time series data by a fixed rate.

5-3- Pre-Processing

After the data collection and data augmentation, pre-processing [3] is the initial phase in training SER. Framing, Windowing, Voice activity detection, Normalization, and Noise reduction are pre-processing steps.

Framing: Speech signals are quasiperiodic and vary over time. Hence, information and related emotions also fluctuate over time. The segregation of the signals happens in a shorter period to make the speech signal invariant. Twenty milliseconds to thirty milliseconds helps to make the speech signal invariant, and five milliseconds of overlap of the frames avoid data leakage between the frames.

Windowing: Windowing on each frame helps to diminish data leakage during Fast Fourier Transformation (FFT) after framing. The Hamming window allows this step where the window size is N for the frames $W(p)$, used in equation (1).

$$W(p) = 0.54 - 0.46\cos\left(\frac{2\pi p}{N-1}\right), \quad for\ 0 \le p \le N-1 \qquad (1)$$

Voice Activity Detection (VAD): An utterance has three portions of speech activities: voiced, unvoiced, and silent. Zero Crossing Rate (ZCR) speech feature helps to detect VAD. ZCR represents the frequency of signal transitions between positive and negative values within a specific frame. Due to high energy, the ZCR value is low for voiced speech and high for unvoiced speech because of low energy.

Normalization: Normalization helps to reduce speaker and recording inconsistency without affecting the features' strength and enhancing the features' generalization capability. The Z-normalization method is used more and represented as

$$Z = \frac{y-\mu}{\sigma} \qquad (2)$$

Where y is the speech signal, μ and σ are the mean and standard deviation of the data, respectively.

Noise Reduction: The environmental noise captured while recording a speech signal affects the recognition rate. Minimum mean square error (MMSE) and log-spectral amplitude MMSE (LogMMSE) reduce the noise from the speech signals. Noise reduction helps to get more accuracy in SER evaluation.

5-4- Feature Extraction

Specific emotions are present in prosodic and spectral features of speech. The extraction of prosodic and spectral features is a vital characteristic of emotion recognition after pre-processing the speech signals.

5-4-1-Prosodic Features

Prosodic features deal with the audio qualities of a speech when connected speeches use sounds as input. The production of the speech deals with the amount of energy, frequency, period, loudness, pitch, and duration. Speech signal communication depends on intonation, stress, and rhythm, which prosodic features can detect. The prosodic features used in the implementation are:

Zero-Crossing Rate (ZCR): ZCR [17] measures the frequency of signal transitions between positive and negative values in every audio frame and defined as

$$Z = \frac{1}{2W_L}\sum_{m=1}^{W_L} sig\,(x_k[m]) - sig\,(x_k[m-1) \qquad (3)$$

Where $sig(.)$ is the sign function.

$$sig[x_k(m)] = \begin{cases} 1, & k \ge 1 \\ -1, & k < 1 \end{cases}$$

Fundamental Frequency (Pitch): Fundamental Frequency (F0) [18] is the minimum frequency of the periodic waveform. F0 is the significant parameter to differentiate male speech from female speech. Pitch also determines the voiced and unvoiced portion of the speech signal. This analysis uses pitch parameters like pitch mean value and pitch range.

The pitch range determines the number of octaves a speech sample can cover, from the lowest to the highest. F0 value varies from 85Hz to 180 HZ for the voiced speech of adult males.

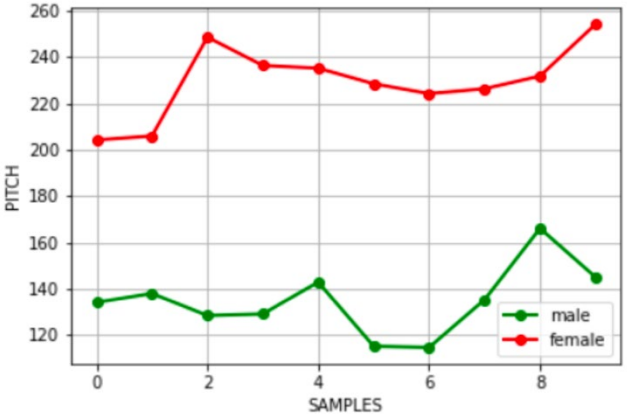


Fig. 4 Representation of Pitch data

Similarly, the F0 value for adult females goes from 165Hz to 255Hz, as shown in Fig 4.

5-4-2-Spectral Features

The representation of spectral features happens by converting the time-domain speech signal into the frequency domain with the help of the Fast Fourier Transform (FFT) which benefits to represent the characteristics of the Human Vocal Tract. The spectral features are

Spectral Centroid (SC): The Spectral Centroid [19] is the most used speech feature where the positioning of the "center of mass" of the audio signal spectrum defines the weight of the spectrum. The center of mass measures the weighted average of the frequency component located in the audio signal, defined as

$$X_{rms} = \frac{\sum_{k=0}^{k-1} f\,(k) y(k)}{\sum_{k=0}^{k-1} y(k)} \quad (4)$$

Where $y(k)$ represents the magnitude of bin number k and $f\,(k)$ represents the central frequency of the bin.

Chroma STFT: Chroma STFT [20] is obtained using FFT on speech samples and the resulting spectrums are a chromatogram in a vertical axis. This feature captures the harmonic feature of the speech signals.

Mel-scale Spectrogram: Mel-scale Spectrogram [21] is a spectrogram in which frequencies convert to the Mel scale, which helps to differentiate the range of frequencies. Mel-spectrogram helps to understand emotions in a better way as humans can perceive sound on a logarithmic scale.

Mel Frequency Cepstral Coefficient (MFCC): MFCC [17] is SER's standard feature extraction technique. The vocal cords, tongue, and teeth filter the sound and make it unique for each speaker in the Human Speech production system. Mel scale represents MFCC, where the frequency bands are equally spaced and close to the Human Auditory System's response, as shown in Fig 5.

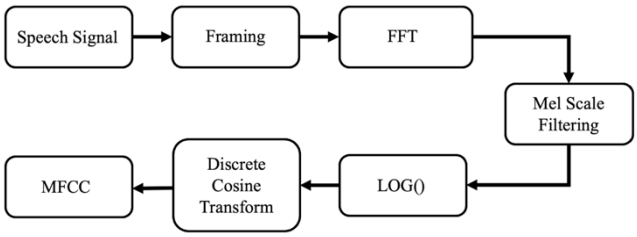


Fig 5 MFCC speech feature extraction

Shifted Delta Cepstrum (SDC): Time derivatives apply to the cepstral coefficients obtained from the MFCC and combined with the delta coefficient to get SDC, as shown in Fig 6. M, D, P and K are the four parameters used in SDC [22]. M denotes the cepstral coefficient for each frame. P indicates added frames. K are the frames that append delta features from the new feature vector. D represents the delta values difference. So, the coefficient vectors represent as

$$c(t) = [c_1, c_2, \ldots\ldots, c_i \ldots\ldots c_{K-1}] \quad (5)$$

c_i are MFCC coefficients and t is the coefficient index. For a given time, t an intermediate calculation is done to obtain the K coefficients.

$$\Delta c_i(t,i) = c(t + i \times P + D) - c(t + i \times P - D) \quad (6)$$

Finally, the SDC coefficients vectors of K dimensions are obtained as

$$SDC(t) = [\Delta c(t,0), \Delta c\big(t, 1 \ldots\ldots\ldots, \Delta c(t, K-1)\big)] \quad (7)$$

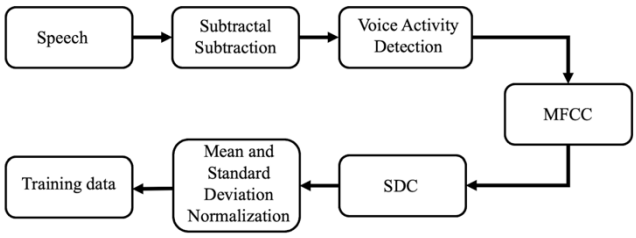


Fig. 6 Extraction of SDC speech feature

This speech feature helps to identify genders for a long-range dynamic appearance in speech signals.

Spectral Flux: Spectral flux [20] measures the change in speed of the signal's power spectrum compared to the previous frame. This feature estimates the speech signal's power spectrum about the power spectrum of one frame with others.

5-5- Feature Selection and Dimension Reduction

Feature selection [23] is selecting features from a large set of extracted features to eliminate redundant and unused information and decrease processing time. For this work, the extracted pitch contains lots of information. The global features, like the range of pitch values and mean pitch values, are selected. Removing redundant and unused information, such as additional zeroes, duplicate values, or frames, is part of MFCC feature extraction. The exact process removes the silent speech frames.

Fundamental frequency and SDC are two speech features to identify gender. Similarly, Spectral centroid, Chroma STFT, MFCC, and Spectral flux features are used for emotion identification to create a single set of data points from multiple speech features using multifeature fusion technique [24].

Multifeature Fusion: The extracted speech features have different dimensions, so feature proximity usages average dimensional spacing between the vectors to denote the proximity between the diverse features [24]. The average dimensional spacing between the vectors is computed as

$$d_{\mu}(k,l) = \frac{1}{N_K N_l} \sqrt{\sum_{N_k} \sum_{N_l} (p_k - p_l)^2} \quad (8)$$

$$d_{\sigma^2} = \frac{1}{N_K N_l} \sqrt{\sum_{N_k} \sum_{N_l} (q_k - q)^2} \quad (9)$$

$d_\mu(k,l)$ and d_{σ^2} are the mean and variance interval of average dimensions of the feature vectors. p_k and p_l are the mean value of k and l type of sound features. q_k and q_l are the mean value of k and l type of sound features. Then the subsequent dimensionality reduction [25] method follows once the feature selection process is complete. High data variance is present in the extracted features containing more information. Dimension reduction techniques reduce the dimensions from extracted feature vectors.

5-5-1- Principal Component Analysis (PCA)

PCA [25] is an approach for reducing the dimensionality of extensive datasets, transforming them into more compact representations while retaining crucial information. These reductions in the number of variables help to get more accurate results. A reduced dataset makes it easier and faster to visualize and analyze the data. Following steps followed to explore PCA.

- **Standardization** creates a normalized dataset when there is a significant difference in the range of initial variables or a larger range of datasets dominates the smaller range. It can be done by
$$Z = \frac{value - mean}{Standard\ Deviation} \quad (10)$$
- **Covariance Matrix Computation** helps to know how the input dataset varies from the mean value to each data point.
- **Eigenvectors and Eigenvalues** of the covariance matrix help to identify the principal component.

5-6- Modelling

This implementation uses two deep learning models. Bidirectional Long Short-Term Memory (BiLSTM) classifies the genders in whispered speech, and Deep Convolutional Neural Network (DCNN) identifies emotions.

BiLSTM [26] combines two Recurrent Neural Networks (RNN) that are placed independently and can traverse backward and forward directions at each time step, as shown in Fig 7. There are two ways to deal with data in this model - The first involves processing data from the past to the future, and the second operates in the opposite direction, from future to past, where two hidden layers of neurons help to preserve the information in both directions. This approach helps to improve the information available in the algorithm.

For BiLSTM, the sequence of inputs is $i = (i_1, i_2, i_3, \dots\dots\dots i_n)$ for a traditional RNN, which computes the hidden vector sequences $h = (h_1, \mathrm{h}, h_3, \dots\dots\dots h_n)$ and results in the output sequences $o = (o_1, o_2, o_3, \dots\dots\dots o_n)$ for the iteration $T = 1$ to n

$$h_T = H(W_{ih} i_T + W_{hn} h_T + b_h) \quad (11)$$

$$o_T = W_{ho} h_T + b_o \quad (12)$$

W are the weight matrices; b is the bias vector, and H is the hidden layer function.

DCNN [27] is a Convolutional Neural Network (CNN) type that helps to identify emotions from whispered speech. The input to the model is the speech data that traverses through many stages, such as the convolution layer, the pooling layer, Activation, Fully Connected, Batch normalization, and dropout layers, as shown in Fig 8.

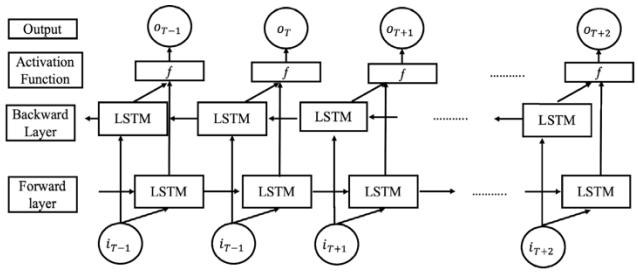


Fig 7. Structure of BiLSTM

To avoid overfitting, Leaky ReLU is used as an activation function in this implementation and evaluated as

$$\begin{cases} i & if\ i \le 0 \\ 0.01\, i & Otherwise \end{cases} \quad (13)$$

The fully connected layer is a loss layer, measuring the inconsistency between the desired and actual outputs. Root The Mean Squared Propagation (RMSProp) optimization algorithm enhances the loss function, which varies in vertical directions.

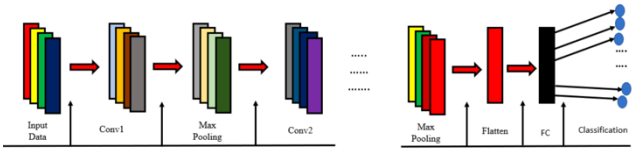


Fig. 8 Structure of Deep Convolution Neural Network

5-7- Emotions

Deep learning techniques are applied to identify the emotions of human speech. Discrete Emotion [3] theory uses seven vital emotions in Human activities.

- **Sadness:** This is the feeling of dissatisfaction, sorrow, or fruitlessness.
- **Happiness:** This is the emotional state that elicits satisfaction and pleasure.
- **Fear:** This feeling triggers a fright response.
- **Anger:** This emotional state leads to feelings of hostility and frustration.
- **Surprise:** It is the state of mind that expresses either positive or negative following something unexpected.
- **Disgust:** A strong emotion that results in feeling repulsed.

- **Neutral:** This is the feeling of lack of particular preference.

5-8- Algorithm for Emotion Recognition

The algorithm to detect emotions from the whispered speech is in two parts

- Gender Detection
- Emotion Identification

Algorithm 1: Geneder Detection

INPUT: Text File (Whispered Speech Samples)
OUTPUT: Emotions (Sad, Happy, Fear, Anger, Surprise, Disgust, Neutral)

1: Begin:
2: Read speech samples from Corpus
3: Pre-processing the speech samples:
4: *Processed_speech=Pre_processing (Speech_Sample)*
5: Extracted SDC, Pitch features:
6: *SDC, Pitch=Feature_Extraction(Processed_speech)*
7: Multifeature fusion:
8: *multi_feature = multifeatured_fusion(SDC, Pitch)*
9: Dimension reduction:
10: *Dimension_R = Dimension_Reduction(multi _feature)*
11: Determine gender:
12: *Gender = BiLSTM (Dimension_R)*
13: Placed in folders:
14: *Gender= (Male, Female)*
15: End:

Algorithm 2 Emotion Identification

1: Begin:
2: Speech Samples = (Male, Female)
3: Data Augmentation:
4: Augemented_samples = Data Augmentation (Speech Samples)
5: Pre-processed the speech samples:
6: Processed_speech=Preprocessing (Augemented_samples)
7: Extracted Spectral Centroid, ChromaSTFT, MFCC, Mel spectrogram features:
8: Spectral_Centroid, Chroma_STFT, MFCC, Mel_spectrogram = Feature Extraction (Processed_speech)
9: Multifeature fusion:
10: multi_feature = multifeatured_fusion(Spectral_Centroid, Chroma_STFT, MFCC, Mel_spectrogram)
11: Dimension reduction:
12: *Dimension_R = Dimension_Reduction(multi_feature)*
13: Emotion detection:
14: *Emotions = DCNN(Dimension_R)*
15: *OUTPUT Emotions (Sad, Happy, Fear, Anger, Surprise, Disgust, Neutral)*
16: End:

6- Experiment and Result Analysis

The proposed SER has been implemented utilizing Python programming language and fortified by various machine learning libraries and additional supporting libraries. The experiment uses Python (Python 3.6.3rc1) and Librosa (Librosa 0.8.0) for audio processing. Graph plotting uses Seaborn and Matplotlib libraries, which help to analyze the speech data. BiLSTM and DCNN models implement the model using Keras (Keras- 2.6.0), TensorFlow (TensorFlow-2.6.0), and Scikit-learn libraries.
wTIMIT dataset is collected and divided into the train, cross-validation, and test samples. Subsequently, the speech processing takes place during implementation, incorporating data-augmentation techniques like shifting time, changing pitch, and changing speed. The pre-processing of speech content from each sample helps to analyze it and extract its features. The dimensionally reduced extracted features align with the data points in the BiLSTM model. The training dataset makes the model learn the features or data. Epochs continuously learn hidden features when the dataset completes backward and forward iterations. Then the cross-validation dataset is used to estimate the model's performance on each epoch and prevent the model from being overfitted. The test dataset helps to evaluate the trained model unbiasedly using performance metrics like precision, recall, f1 score, and confusion Matrix.
The parameters used in BiLSTM models are as in Table 1.

Table 1: Parameters of the BiLSTM model

Layers	***Shape of output***	***Parameters#***
Embedding (Embedding)	(None, 216, 256)	524544
Bidirectional	(None, 128)	98816
batch normalization	(None, 128)	512
Dense (Dense)	(None, 128)	10512
dropout (Dropout)	(None, 128)	0
flatten (Flatten)	(None, 128)	1024
dense1(Dense)	(None, 12)	0
dense2(Dense)	(None, 12)	455

The BiLSTM model helps to segregate the speech samples into male and female. The segregated speech samples move to their respective male and female folders.

Table 2: Parameters of the DCNN model

Layers	***Shape of output***	***Parameters#***
conv1d (Conv1D)	(None, 216, 256)	2304
Activation (Activation)	(None, 216, 256)	0
conv1d 1 (Conv1D)	(None, 216, 256)	524544
batch normalization	(None, 216, 256)	1024
activation 1 (Activation)	(None, 216, 256)	0
dropout (Dropout)	(None, 216, 256)	0
max pooling1d (MaxPooling1D)	(None, 13, 256)	0
conv1d 2 (Conv1D)	(None, 13, 128)	262272
activation 2 (Activation)	(None, 13, 128)	0
conv1d 3 (Conv1D)	(None, 13, 128)	131200
activation 3 (Activation)	(None, 13, 128)	0
conv1d 4 (Conv1D)	(None, 13, 128)	131200
batch normalization	(None, 13, 128)	512
activation 4 (Activation)	(None, 13, 128)	0
conv1d 5 (Conv1D)	(None, 13, 128)	131200
batch normalization 1	(None, 13, 128)	512
activation 5 (Activation)	(None, 13, 128)	0
dropout 1 (Dropout)	(None, 13, 128)	0
max pooling1d 1 (MaxPooling1D)	(None, 1, 128)	0
conv1d 6 (Conv1D)	(None, 1, 64)	65600
activation 6 (Activation)	(None, 1, 64)	0
conv1d 7 (Conv1D)	(None, 1, 64)	32832
activation 7 (Activation)	(None, 1, 64)	0
flatten (Flatten)	(None, 64)	0
dense (Dense)	(None, 14)	910
activation 8 (Activation)	(None, 14)	0

After separating the speech samples into genders, Chroma STFT, Spectral centroid, Mel-scale spectrogram, and Spectral Flux features extraction happened. Then the created data points are dimensionally reduced to fit into the DCNN deep learning model to identify emotions. Metrics like precision, recall, F1 score, and the confusion matrix aid in assessing the model's performance. Table 2 mentions the parameters of the DCNN model.

6-1- Result Analysis

The fusion of Fundamental frequency and SDC detects the gender of speech samples. The initial stage involves pre-processing the speech samples, followed by feature extraction, which includes SDC and fundamental frequencies. The combination of SDC and Fundamental speech features create a multifeature fusion resulting in a single set of data points. The next step is to reduce the dimensions of data points to use them as input to the BiLSTM Model. The BiLSTM model predicts the data in the dataset as male and female, placed in separate folders. Of these, 3342 speech samples are available, and 3294 are correctly classified. So, the accuracy of the model prediction is 99.59%. The precision, recall, and f1-score values, along with the confusion matrix, are shown in Figs 9 and 10. The predicted model identifies the female speech samples more accurately than the male.

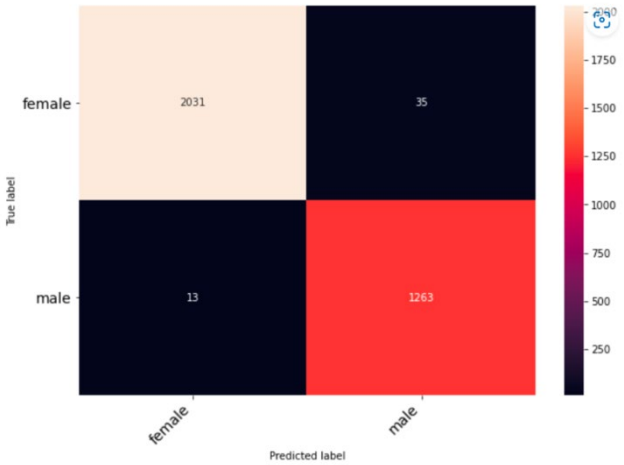


Fig. 9 Confusion Matrix for Gender detection by BiLSTM.

The graph in Fig 9 shows the gender detection from the speech sample. The X-axis represents the true label, and the Y-axis shows the predicted label. This graph gives the

count of detected gender speech that the Bi-LSTM model correctly and incorrectly detects.

	precision	recall	f1-score	support
female	0.99	0.98	0.99	2066
male	0.97	0.99	0.98	1276
accuracy			0.99	3342
macro avg	0.98	0.99	0.98	3342
weighted avg	0.99	0.99	0.99	3342

Fig. 10 Accuracy measures of Gender Detection.

2031 female and 1263 male speech are correctly detected, whereas 13 females and 35 males are incorrectly detected. Pitch values identify the emotions after detecting genders from the speech samples.

The extraction of MFCC, Mel-scale spectrogram, Chroma STFT, Spectral Flux, and Spectral Centroid speech features happens from speech samples. Then, all five speech features are fused into a single feature and dimensionally reduced. Finally, the DCNN model is applied to predict emotions.

Individual emotions are detected based on gender as shown in Fig 11. The deviation in the male speech emotions is less than the female. Female fear and female neutral emotions show divergent results from other emotions. The accuracy of the model is 98.54%.

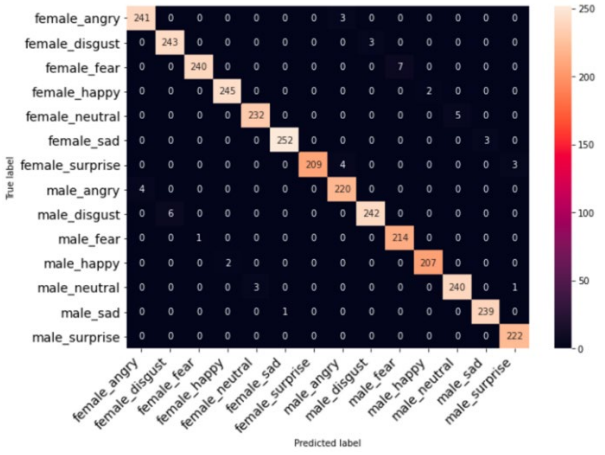


Fig.11 Emotions based on gender.

6-2- Comparison Analysis

This experiment evaluates the current implementation using the wTIMIT dataset, resulting in an impressive accuracy of 98.54%. The performance compares with three-layered Long short-term memory Bidirectional RNNs (LSTM BiRNN) that use No-Attention, Feed Forward Attention (FFA), and Improved Feed Forward Attention Mechanism (IFFA) to evaluate the emotions in the dataset. This model uses 35 hidden states to train the dataset. The models are trained and validated with the help of the wTIMIT dataset [28] and tested with the help of the CHAINS dataset, as shown in Table 3. This experiment is a new hypothesis for emotion detection with gender identification on whispered speech. Hence, there are fewer references available for similar investigations.

Table 3: Comparison of the Implementation

Sl#	***Model Implementation***	***Accuracy***
1	FFA, IFFA, LSTM, Bi_RNN [28]	97.6 %
2	Proposed Model	98.54 %

The proposed method improves the performance of emotions after segregating the speech samples into genders. By separating the approach into two models, this implementation excels in capturing natural and spontaneous expressions of emotions with low latency manner.

7- Conclusion and Future Work

Identification of gender from whispered speech and recognizing emotions is a complicated task. This implementation initially detects the gender from the speech samples before identifying the emotions. Emotions might vary based on gender in the same situation. Speech features such as Fundamental Frequency and SDC help with gender identification. MFCC, Mel-scale spectrogram, Spectral flux, Spectral Centroid, and Chroma STFT play a vital role in detecting emotions. Multifeature fusion helps to combine speech features into a single set of data points. The concept verifies with the publicly available dataset wTIMIT with an accuracy of 98.54%. This approach helps to identify nearly inaudible emotions and is used to figure out the strategy. The proposed methodology can be improved using other speech features, machine learning, and deep learning concepts.

References

[1] ST Jovicic, and Z Saric, “Acoustic analysis of consonants in whispered speech,” Journal of voice, vol 22, no. 3, pp. 263–274, 2008.

[2] M Kumari, and I Ali, “An efficient algorithm for gender detection using voice samples,” 2015 Communication, Control and Intelligent Systems (CCIS), 2015 , Mathura, Utter Pradesh, pp. 221–226, doi: 10.1109/CCIntelS.2015.7437912.

[3] S Motamed, S Setayeshi, A Rabiee, and A Sharifi, “Speech Emotion Recognition Based on Fusion Method,” Journal of Information Systems and Telecommunication (JIST), vol. 3, pp. 50--56, 2017, doi: 10.7508/jist.2017.17.007.

[4] JS Li, CC Huang, ST Sheu, and MW Lin, “Speech emotion recognition and its applications,” Proc. of Taiwan Institute of Kansei Conference, 2010 Paris, France, pp. 187–192.

[5] A Guerrieri, E Braccili, F Sgro, and GN Meldolesi "Gender identification in a two-level hierarchical speech emotion recognition system for an Italian Social Robot," Sensors, vol. 22, no. 5, pp. 1714, 2022, doi: 10.3390/s22051714.

[6] M Sarria-Paja, TH Falk, and D O'Shaughnessy, "Whispered speaker verification and gender detection using weighted instantaneous frequencies," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013 (May 26-31), Vancouver, Canada, pp. 7209–7213, doi: 10.1109/ICASSP.2013.6639062.

[7] J Deng, S Fruhholz, Z Zhang, and Bojrn Schuller, "Recognizing emotions from whispered speech based on acoustic feature transfer learning," IEEE Access, vol. 5, pp. 5235–5246, 2017.

[8] M Cotescu, T Drugman, G Huybrechts, J Lorenzo-Trueba, and A Moinet, "Voice conversion for whispered speech synthesis," IEEE Signal Processing Letters, vol. 27, pp. 186–190, 2019, doi: 10.1109/LSP.2019.2961213.

[9] P Mishra, and R Sharma, "Gender differentiated convolutional neural networks for speech emotion recognition," 2020 12th International Congress on Ultra-Modern Telecommuni- cations and Control Systems and Workshops (ICUMT), 2020 (October 5-7), Brno, Czech Republic, pp. 142–148, doi: 10.1109/ICUMT51630.2020.9222412.

[10] Mustaqeem S Kwon, "Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network," International Journal of Intelligent Systems, vol. 36, no. 9, pp. 5116– 5135, 2021, doi: 10.1002/int.22505.

[11] J. Ancilin and A. Milton, "Improved speech emotion recognition with mel frequency magnitude coefficient," Applied Acoustics, vol. 179, pp. 108046, 2021, doi: 10.1016/j.apacoust.2021.108046.

[12] S. Jothimani and K. Premalatha, "Mff-saug: Multi feature fusion with spectrogram augmentation of speech emotion recognition using convolution neural network," Chaos, Solitons & Fractals, vol. 162, pp. 112512, 2022, doi: 10.1016/j.chaos.2022.112512.

[13] B Yalamanchili, SK Samayamantula, and KR Anne, "Neural network-based blended ensemble learning for speech emotion recognition," Multidimensional Systems and Signal Processing, vol. 33, no. 4, pp. 1323--1348, 2022, doi: 10.1007/s11045-022-00845-9.

[14] T Feng, R Hebbar, and S Narayanan, "Trustser: On the trustworthiness of fine-tuning pre-trained speech embeddings for speech emotion recognition," arXiv preprint arXiv:2305.11229, 2023, doi: 10.48550/arXiv.2305.11229

[15] RV Darekar and M Chavan, S Sharanyaa, and NR Ranjan, "A hybrid meta-heuristic ensemble based classification technique speech emotion recognition," Advances in Engineering Software, vol. 180, pp. 103412, 2023.

[16] J Rekimoto, "Dualvoice: A speech interaction method using whisper-voice as commands," CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1–6, 2022, doi: 10.1145/3491101.3519700.

[17]H Dolka, AX VM, and S Juliet," Speech emotion recognition using ANN on MFCC features," 2021 3rd international conference on signal processing and communication (ICPSC), 2021, Coimbatore, India, pp. 431–435, doi: 10.1109/ICSPC51351.2021.9451810.

[18] MK Reddy and KS Rao, "Robust pitch extraction method for the hmm-based speech synthesis system," IEEE signal processing letters, vol. 24, no. 8, pp. 1133–1137, 2017, doi: 10.1109/LSP.2017.2712646.

[19] J Chatterjee, V Mukesh, HH Hsu, G Vyas, and Z Liu, "Speech emotion recognition using cross- correlation and acoustic features," 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, 2018 (August 12-15), Athens, Greece, pp. 243–249, doi: 10.1109/DASC/PiCom/DataCom/CyberSci Tec.2018.00050.

[20] S Rajesh and NJ Nalini, "Musical instrument emotion recognition using deep recurrent neural network," Procedia Computer Science, vol. 167, pp. 16--25, 2020, doi: 10.1016/j.procs.2020.03.178.

[21] M Aly, and NS Alotaibi, "A novel deep learning model to detect covid-19 based on wavelet features extracted from mel- scale spectrogram of patients' cough and breathing sounds," Informatics in Medicine Unlocked, vol. 32, pp. 101049, 2022, doi: 10.1016/j.imu.2022.101049.

[22] Z Qawaqneh, AA Mallouh, and BD Barkana, "Age and gender classification from speech and face images by jointly fine-tuned deep neural networks," Expert Systems with Applications, vol. 85, pp. 76–86, 2017, doi: 10.1016/j.eswa.2017.05.037.

[23] A Koduru, HB Valiveti and AK Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," International Journal of Speech Technology, vol. 23, no. 1, pp. 45–55, 2020, doi: 10.1007/s10772-020-09672-4.

[24] S Zhang, and C Li, "Research on feature fusion speech emotion recognition technology for smart teaching," Mobile Information Systems, vol. 2022, 2022, doi: 10.1155/2022/7785929.

[25] GB, Prasanna, SV Bhat, C Naik, and HN Champa, "An Efficient Method for Handwritten Kannada Digit Recognition based on PCA and SVM Classifier," Journal of Information Systems and Telecommunication (JIST), vol. 3, no. 35, pp. 169 2021, doi: 20.1001.1.23221437.2021.9.35.3.2.

[26] A Graves, N Jaitly, and A Mohamed, "Hybrid speech recognition with deep bidirectional lstm," 2013 IEEE workshop on automatic speech recognition and understanding, 2013, Olomouc, Czech Republic, pp. 273–278, doi: 10.1109/ASRU.2013.6707742.

[27] N Aloysius and M Geetha, "A review on deep convolutional neural networks," 2017 international conference on communication and signal processing (ICCSP), 2017, pp. 0588–0592, IEEE, doi: 10.1109/ICCSP.2017.8286426.

[28] SBC Gutha, MAB Shaik, T Udayakumar, and AA Saunshikhar, "Improved feed forward attention mechanism in bidirectional recurrent neural networks for robust sequence classification," 2020 International Conference on Signal Processing and Communications (SPCOM), 2020, IISc, Bangalore. IEEE, pp. 1—5, doi: 0.917960610.1109/SPCOM50965.202

Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 162-169
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Ensemble learning of Ada-boosting Based on Deep Weighting for Classification of Hand-written Numbers in Persian (With the doctors' prescription approach)

Amir Asil[1], Hamed Alipour[2*], Shahram Mojtahedzadeh[1], Hasan Asil[3]

[1].Department of Electrical Engineering, Faculty of Electrical and Computer Engineering, Azarshahr Branch, Islamic Azad University, Iran
[2] .Department of Electrical Engineering, Faculty of Engineering, Tabriz Branch, Islamic Azad University, Iran
[3] .Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Azarshahr Branch, Islamic Azad University, Iran

Received: 03 Feb 2023/ Revised: 02 Feb 2024/ Accepted: 06 Jun 2024

Abstract

Converting handwritten data to electronic data is one of the challenges that have been raised over the past years. Considering that these data are used in various sciences, solving this challenge is of great importance. One of these sciences is medical science that doctors use in prescriptions. This project tries to classify handwritten numbers with the approach of solving the challenges of handwritten data. Over the past years, a variety of solutions have been developed to transform handwritten data based on machine learning. Each method categorizes or clusters the data based on the type of data and its use. In this paper, a new approach based on hybrid methods and deep learning is presented for the classification of Persian handwritten data. By combining Ada and convolution, a deeper examination of the data is performed in basic learning. The purpose of this research is to provide a new technique for classifying images of Persian handwritten numbers. The structure of this technique is based on Ada Boosting, which in turn is based on weak learning. This technique improves learning by repeating weak learning processes and updating weights. Meanwhile, the proposed method tried to employ stronger language learners and provide a stronger algorithm by combining these strong learners. This method was evaluated on the Hoda standard dataset containing 60,000 training data. The results show that the proposed method has more than 1% less error than the previous methods. In the future, as the base learner develops, new mechanisms can be introduced to improve results with new types of learning.

Keywords: Deep Learning; Ada Boosting; Handwritten Data; Convolution; Classification.

1- Introduction

In the modern information age, the volume of electronic documents and files are is huge [1]. Processing of these files requires lot of time and cost. For example, there are great volumes of handwritten data that cannot be converted to electronic text manually [2]. In recent years, the importance of using high speed computer systems has been taken into account. Different approaches are used to infer knowledge from such documents. Classification is one of these methods. Classification and clustering are the most important tools in artificial intelligence [3].

In artificial intelligence, deep learning is a machine learning technique that teaches computers to learn what human beings do naturally.

In deep learning, a computer model learns how to classify images, text or sound directly. Deep learning models are highly accurate and sometimes their performance exceeds that of the human beings. Deep learning methods have become very popular in recent years and were used in various projects [4]. In deep learning, the nonlinear properties of several layers are extracted and transferred to a classifier and a combining layer to combine features and make predictions [5]. The deeper the hierarchy of the layers, the more nonlinear properties are obtained and the better the results. Convolution, DBNs, etc. are examples of

✉ **Hamed Alipour**
Hamed.Alipour@iaut.ac.ir

these networks that have many applications such as image processing [6].

In addition to various deep learning techniques, combinatorial approaches to machine learning have also been considered. Combined methods are general techniques of machine learning. By combining different predictions, these methods try to provide more accurate results in solving problems. These include boosting, bagging, and stacking [7].

The present aims at combining deep learning methods with combining methods. This method aims to reduce the error rate in the classification of Persian letters with by using the proposed technique. For this combination in this study, adaboosting is combined with convolution.

2- Literature Review

Only few studies have been performed on the classification of Persian handwritten letters based on machine learning. One of these studies has been on recognition of handwritten numbers based on pre-classification [8]. In another study, Mahabadi et al. used a fuzzy method for the classification of images [9]. Soltanzadeh et al. used the gradient and support vector to classify Persian images [10]. In another study, an intelligent method was presented for feature selection based on binary gravitational search algorithm in Persian handwritten number recognition system. In this method, the fitness function associated with Persian handwritten recognition system error is minimized by using binary gravitational search algorithm and by selecting appropriate features. Implementation results show that the intelligent feature selection technique is able to select the most effective features for the recognition system [9]. In another study, rotation-invariant classification was introduced [10]. In another research, a new method called NeuroWrite has been presented. In this unique method, deep neural networks have been used to predict the classification of handwritten digits [11]. This model is extremely accurate in identifying and classifying handwritten figures using the power of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). As you know, recognition of handwritten digits is a topic of interest for computer vision scientists [12,13]. In the research based on this method, a satisfactory and appropriate algorithm for this multi-class classification problem (0-9) has been presented. The purpose of this research is to compare seven machine learning algorithms in terms of their performance criteria in recognizing handwritten digits using two datasets [14,15]. In this research, models of nearest neighbors (kNN), support vector machine (SVM), logistic regression, neural network, random forest (RF), simple Bayes and decision tree based on this in relation to the area under the curve (AUC), accuracy (ACC)) are evaluated [16]. The National Institute of Standards and Database Technology (MNIST) modified common dataset and the Hand Digit Classification (HDC) dataset were the image providers on which this research was conducted [17]. The results confirm that the neural network model is an excellent classifier for this problem. However, it provides similar results to other machine learning classifiers in several cases. In another research, to capture segment-level sentiment fluctuations in an utterance, sentiment profiles (SPs) have been proposed to express segment-level soft labels [18].

There are two main approaches to combinatorial approaches: including Bagging and Boosting. Extensive research has been done on these two techniques. G-Boosting, AdaBoost, LogitBoost, and GentleBoost ... are some of the subset algorithms of boosting [19, 20].

AdaBoost is one of the first boosting algorithms introduced by Freund and Schapire [21]. The main idea of Boosting Algorithm is to give more weight to samples that have been incorrectly classified by current hypotheses [22, 23]. Various solutions such as classifying into two other classes, etc., have been proposed for the multi-class classification [24, 25]. These methods have been increasingly developed. Another method for multi-class classification is the N-ary technique, which can solve multi-class classification problems by splitting them into binary sets [26]. There is another method based on annotation and tagging of untagged data [27]. Another study was performed on combining methods and deep learning in object recognition [28].

In research conducted by Farkhi and etc, the recognition of Persian handwritten digits was used to read check amounts and postal code digits, etc. The purpose of presenting this research was to identify Persian handwritten digits that are written by different people's handwriting so that they can be used in automation software such as postal code reading in the post office department. To perform and test, a database of handwritten figures is needed, which is available in the mnist database for the English language and the Hoda database for the Farsi language, which had about two thousand samples for each digit. In this research, deep learning has been used to identify Persian handwritten digits, which is implemented in such a way that it is done with a deep Boltzmann machine and a two-layer automatic encoder, in which 200 neurons are used in each layer. The method proposed by this team showed a number recognition percentage of up to ninety-two percent [29].

In machine learning, learning transfer and generalization are important and fundamental capabilities. In the research conducted by Nowrozi, self-supervised learning was used to classify images. In this research, as a monitoring method, by using jigsaw puzzle and guessing the angle of rotation, it has been tried to classify handwritten images by

generalizing the domain. This method has reduced errors and improved [30].

The goal of transfer learning (TL) with convolutional neural networks is to improve performance on a new task using knowledge of similar tasks previously learned. This has greatly helped image analysis as it overcomes the problem of data scarcity and also saves time and hardware resources. However, transfer learning is arbitrarily configured in most studies. In this research, an attempt has been made to provide a solution for choosing the model and TL approaches for image classification work [31].

Extracting information from text images identified from the Internet channel is one of the most important problems of information collection systems in the field of information technology. This problem becomes more acute when we know that among the multitude of text images, only a small percentage of identified text images have informational value. In another research, a classification method based on image zoning was used to analyze text images and access their content. In this algorithm, with the help of a two-step zoning method, the image areas are identified, then with the help of a hierarchical classification structure, the type of the area is determined in terms of text or photo (non-text). In the following, by defining the value of the text of a text image, we try to categorize the text image into one of two semantic groups, valuable and worthless. The proposed algorithm is evaluated on a database of textual and non-textual images provided from images available on the Internet. The results of the tests show the effectiveness of the proposed method in the semantic classification of images based on the user's definition of valuable and worthless textual images. The presented algorithm has provided 98.8% classification accuracy for classifying valuable text images from worthless ones [32].

Local binary pattern is a widely used descriptor in feature extraction from texture images. Convolutional deep neural networks are also considered to be the best classification tools with very high accuracy. In another research, a structure for combining the features of local binary pattern and deep convolutional neural network for the classification of noisy texture images has been presented, which provides very high accuracy for the classification of noisy texture images. This method consists of two feature extraction tools. In one tool, local features of texture images are extracted in the form of a 3D histogram using the complete local binary pattern. In the second tool, texture features are reduced using DenseNet-121 deep convolutional neural network. This part, which is used in the feature combination process, significantly reduces the dimensions of the 3D histogram by using a shallow convolutional neural network to combine with deep features. The accuracy of the proposed model has been evaluated on Outex, CUReT and UIUC noise datasets with Gaussian noise, point noise and salt pepper noise with different intensities, and the classification accuracy of the proposed method for different amounts of noise is improved between 3 has had 15 percent [33].

In another research that has been done on the classification of images using deep neural networks, which has led to the emergence of attention-based classification systems. Unlike the methods that use only one classifier to categorize images, in this research, a method for simultaneously using attention-based classifiers with different attentions and calculating the result of their results has been presented. In this research, two general methods are proposed to calculate the result of the classification results: simple voting and calculation of the result based on Bayes logic. Examining the results of this method on the CIFAR10 data set shows the positive effect of the proposed methods on improving the classification accuracy [34].

Since the classification process is completely dependent on the extracted features, it is necessary to act very intelligently in the extraction and selection of images to achieve the ideal accuracy. In the research conducted by Babaian and etc, the evaluation of deep learning in image classification has been considered. In this research, the Python programming language was implemented using the pytorch library. In the classification of images, 38 categories of objects in two groups of test and training and a total of 2,746 photos were examined. In order to evaluate two methods, deep neural network and support vector machine were compared [35].

Deep learning networks have been used to solve various problems such as image classification, object identification, image extraction, etc [36, 37]. These networks try to optimize these techniques by applying different methods.

The deep models of image classification include two deep models of PixelRNN and DCGAN. A new combined model was also provided by Daniel Fritis based on the combination of PixelRNN and DCGAN to detect images [38]. The Gan is another project for image classification, which is related to K-class classification [39, 40].

By creating different layers, deep networks model data features in more details. They have a have great power [33, 41]. One of these deep networks is the Eight Network, which is based on convolution and has eight layers. [42, 43] This model has been used in a variety of problems including video classification, face recognition and action recognition [44]. Other projects in this area have been the use of convolution as a weak learner in the combined Boosting algorithm [45]. The present study is to present a multi-class model based on the ada boosting and convolution combined method for the classification of Persian images (handwritten numbers).

3- Proposed Algorithm

The present study aims at providing a new technique for classification of the images of handwritten Persian numbers. The structure of this technique is founded on Ada Boosting, which in turn, is based on weak learning. This technique improves learning by iteration of the weak learning processes and updating weights. In the meantime, the proposed method tried to employ stronger learners and present a stronger algorithm by combining these strong learners.

The proposed method is to improve image classification and to reduce errors by combining Ada Boosting and basic learning. Figure 1 shows the general structure of the Ada Boosting method.

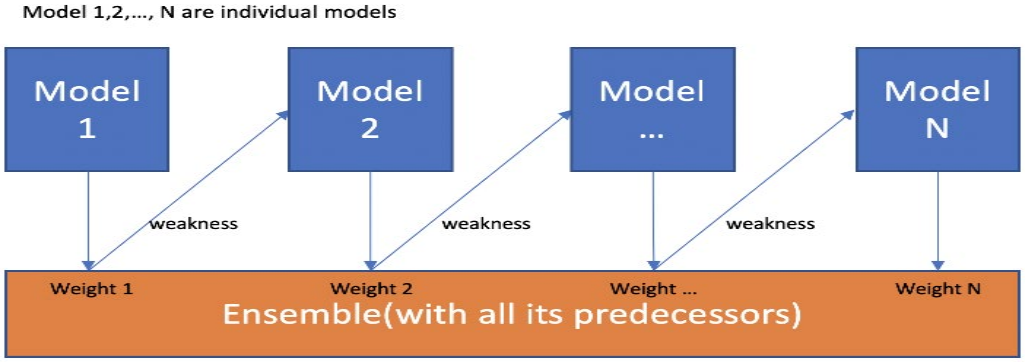


Fig. 1. The structure of the Ada Boosting method [13]

Ada Boosting involves weak learning, and tried to enhance learning by weighting. Proper selection of basic algorithms can help improve learning. Different basic learning methods were used in different problems. This study is to use deep learning to improve this algorithm.

Each hybrid method includes four components of the training set, basic learner, generator, and compiler. The components of the proposed method are:

Training set: A training set includes labeled examples used for training. The training set of this project consists of handwritten data. Some of the data will be used for training and some for testing

Basic Learner: A basic learner is a learning algorithm used to learn a training set. In this algorithm we will use deep learners for the classification. The learning algorithm used in research is convolution, which has two convolution functions for each class. This is in line with greater convergence. On the other hand, the strategy of convolution is reducing the error rate in learning.

Generator: Generator is used to create different classifiers. Different classifiers are created at each step of the Ada boosting method.

Compiler: the compiler is used to combine classification methods. Various methods have been proposed to combine the classifiers. Majority voting is one of the most widely used methods, which functions similar to the non-weighted averaging.

However, rather than averaging out the output probability, the … (?) array counts the predicted labels of the basic learners and makes a final prediction using the highest-rated label. The majority voting can take a non-weighted average using the basic learner label and choose the label with the most value. One of the disadvantages of majority voting is the loss of data because it only uses the predicted labels. Figure 2 shows the general structure of the proposed algorithm.

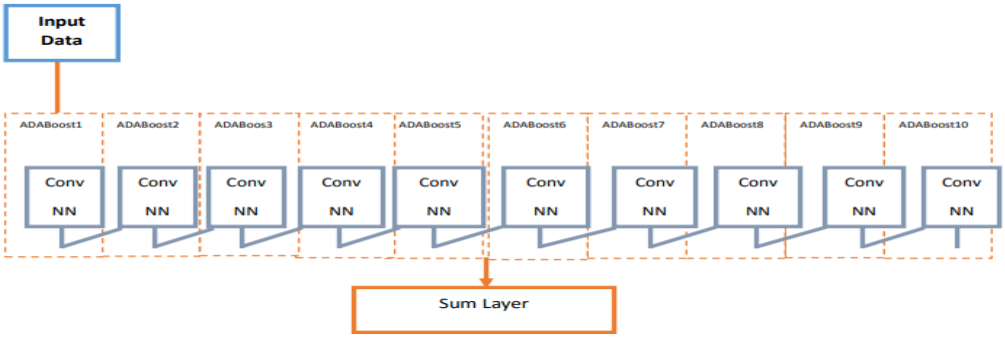


Fig. 2. The structure of the deep Ada boosting method

Let the training sets of (x1, c1), (x2, c2), (x3, c3), and so on; where the input Xi ∈ Rp is the qualitative or quantitative output ci, but it is assumed to be in the set {1,2,3, ⋯ k}. K is the number of classification classes. Training datasets are independent from one another.

The classification aims at finding C(x) based on the training data. With the new input x, we can obtain a class label of {1,. . . , K}.

Ada Boost adopts an iterative procedure. This method predicts the classifier based on the combination of weak learners. At first there are equal weights among the sample. The weights are then updated with each iteration and change in the subsequent classes. The number of iterations in learning is usually between 500 and 1000. The score of each step is taken as a coefficient and multiplied by the classifier. Finally, these classifiers are combined linearly. ALG. 1 shows the Ada Boosting algorithm [26]:

It should be noted that theoretically, the error rate err(m) is less than ½ in each weak classifier [8]. When the number of classes is k, the error probability is (K−1)/K. In Ada-Boosting method, when the error rate is greater than ½, weak learners do not have a high efficiency [26].

Algorithm 1. AdaBoost

1. Initialize the observation weights $w_i = 1/n$ *i = 1, 2,…,n.*

2. For m = 1 to M:

(a) Fit a classifier $T^{(m)}(x)$ *to the training data using weights* w_i.

(b) Compute

$$err^{(m)} = \sum_{n}^{i=1} w_i \prod(c_i \neq t^{(m)}(x_i)) / \sum_{n}^{1=1} w_i$$

(c) Compute

$$\alpha^{(m)} = log \frac{1 - err^{(m)}}{err^{(m)}}$$

(d) Set

$$w_i \leftarrow w_i \,.\exp(\alpha^m \,.\, \prod\left(c_i \neq t^{(m)}(x_i)\right))$$

for i = 1, 2,…,n.

(e) *Re-normalize* w_i.

3. Output C(x)

$$C(x) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \prod_i \left(T^{(m)}(x) = k\right))$$

ALG2. Ada boosting algorithm

This study is to propose a new algorithm based on Ada Boosting to classify multi-classes. We use weak deep learning methods to solve this problem. As mentioned, the two classes of the convolution network are used in the weak learner for classification. Generally, a convolutional neural network is a hierarchical neural network with its convolutional layers adopted alternately with the pooling layers and thereafter, there are a number of interconnected layers. It has high capabilities in partial learning due to the deepness of these networks. This method is to increase the network's capabilities to the optimized level for the Ada boosting. The Ada Boosting method based on weight updating tries to increase the weight of false guesses in successive iterations, so that a better training takes place in the subsequent learning. If we let the weight of weak learning (convolution) to be W' I, Alg. 2 will be the proposed algorithm for this method:

Algorithm 2. Deep AdaBoosting

1. Initialize the observation weights $w_i = 1/n$ *i = 1, 2,…, n.*

2. For m = 1 to M:

(a) Fit a classifier $T^{(m)}(x)$ *to the training data using weights* w_i.

(b) Compute

$$err^{(m)} = \sum_{n}^{i=1} w_i \prod(c_i \neq t^{(m)}(x_i)) / \sum_{n}^{1=1} w_i$$

(c) Compute

$$\alpha^{(m)} = log \frac{1 - err^{(m)}}{err^{(m)}}$$

(d) Set

$$w_i \leftarrow w_i' + (w_i \,.\exp\left(\alpha^m \,.\, \prod\left(c_i \neq t^{(m)}(x_i)\right)\right) + \log(k - 1))/2 \qquad w_i' \text{ is weight CNN in week leaner.}$$

for i = 1, 2,…,n.

(e) *Re-normalize* w_i.

3. Output C(x)

$$C(x) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \prod_i \left(T^{(m)}(x) = k\right))$$

ALG. 2. Proposed algorithm based on combining weights

4- Evaluation of the Proposed Algorithm

In order to evaluate this method, the above algorithm was implemented in MATLAB and evaluated on the Hoda database. Hoda handwritten numbers set is the first large set of Persian handwritten numbers, consisting of 102353 black-and-white handwritten samples. The set was developed in a master's project on handwritten form recognition. The data of this set were collected from about 12000 completed forms.

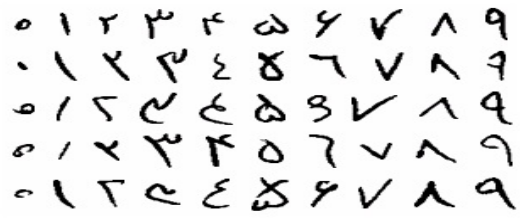


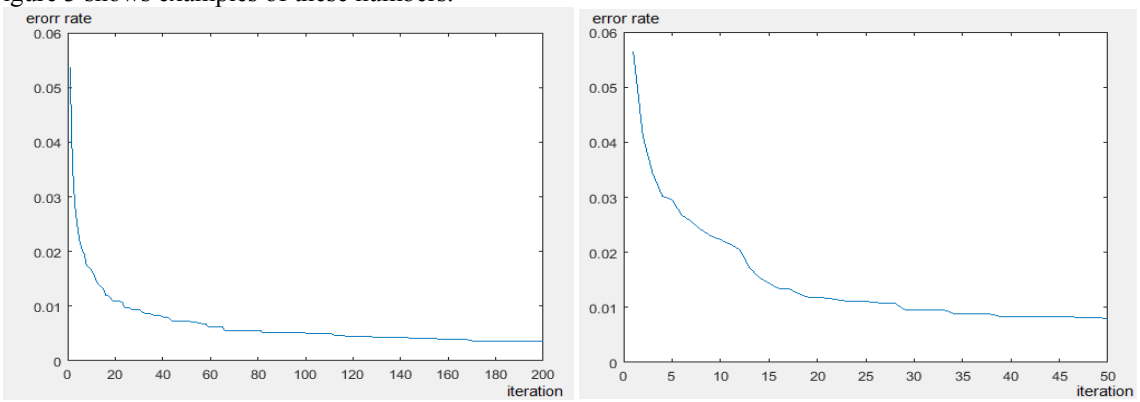
Fig. 3. An example of handwritten data

The number of training samples is 60000 and test samples 20000. Figure 3 shows examples of these numbers.

Previously proposed techniques [30, 31] are used to compare the results of this study. On the other hand, for a more accurately evaluation the results of each test were assessed with a number of iterations. The results of these evaluations are presented in the following section.

Comparison of the results

To evaluate the results of this study, as presented in Section 4, iterative training was performed and tests were made thereafter. Figure 4 shows the evaluation results with a number of iterations.



50 iterations 200 iterations

Fig. 4. Error rate with various numbers of iterations

As shown in Figure 4, the error rate decreases with increasing number of iterations.

error rate: This is a measure of how wrong the classifier would be if it predicted just the majority class. The formula is (Actual: No/Total Sample).

The results are then compared with the literature [30]. Compared with other proposed methods, the error rate is reduced in this method. Table 1 shows the rate of correct classifications.

Table 1: Comparison of error rates by types of algorithms

Deep AdaBoost	*conscious methods*	*different methods*
99.9	94.91	98.16

Figure 5 shows the comparison diagram for the proposed algorithm and different methods. The aim of this project is to reduce the error rate in the classification of handwritten data. After the implementation and training and homogenization of the error rate in the number of repetitions, a comparison has been made.

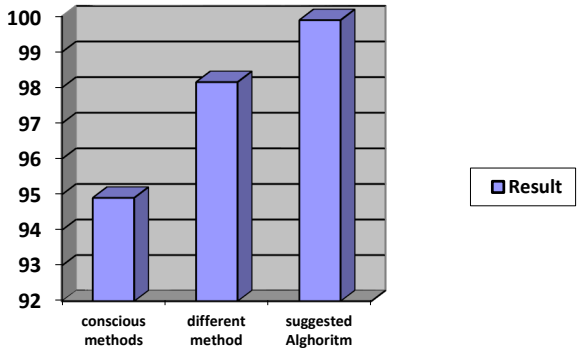


Fig. 5. Results of the correct classification of different algorithms

According to the above, the error rate has reduced more than 1% in the proposed method.

5- Conclusion

Various methods have been proposed for handwritten number classification. But most of these methods were related to English datasets and little was done on Persian datasets. Deep learning methods, especially the convolution technique, have also been used extensively in solving complicated problem and in image processing.

Convolution is one of the learning methods offered for images classification. In fact, combinatorial classifications are presented to enhance the network capabilities. In this research, a combination of Ada Boosting and convolution methods was adopted to present a new model for the classification of Persian handwritten data. Due to the advantages of the convolution method and the Ada Boosting, the combinational updating of weights was performed in the proposed algorithm. The results show that the proposed method has a lower error rate than the previous methods by more than 1%. In the future, by developing basic learner, new mechanisms can be provided to improve the results by new types of learning.

References

[1] Maziyar Kazemi, Muhammad Yousefnezhad, Saber Nourian, "Persian Handwritten Letter Recognition Using Ensemble SVM Classifiers Based on Feature Extraction", National Conference on Intelligent Systems and Information and Communications Technology, At Tabriz, Iran, Volume: 12

[2] Addakiri K., Bahaj M. (2012) "On-line Handwritten Arabic Character Recognition using Artificial Neural Network", International Journal of Computer Applications (IJCA), Volume 55.

[3] Alizadeh H., Yous efnezhad M., Minaei-Bidgoli B. (2015)"Wisdom of Crowds Cluster Ensemble", Intelligent Data Analysis, IOS Press, Vol. 19(3).

[4] sadafsafavi, mehrdad jalali, Recommendation model of places of interest according to people's behavior pattern based on friends list based on deep learning, Journal of Information Systems and Telecommunication (JIST),2021-11-17,http://ijece.org/en/Article/29095

[5] Jamshid bagherzadeh, hasan asil, proposing a New Method of Image Classification Based on the AdaBoost Deep Belief Network Hybrid Method, TELKOMNIKA, Vol 17, No 5, 2019

[6] Mohammad Ebrahim Khademi, Mohammad Fakhredanesh, Farsi Conceptual Text Summarizer: A New Model in Continuous Vector Space, Journal of Information Systems and Telecommunication (JIST),2019-11-04,http://jist.ir/en/Articl /15222

[7] Cheng Ju and Aur´elien Bibaut and Mark J. van der Laan, The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image Classification, cornell University,2017: arXiv:1704.01664v1 [stat.ML] 5 Apr 2017

[8] amin Ollah Mah Abadi, Abdolmajid Jazemian," Fuzzy diagnosis of Persian handwritten numbers" ،THE CSI JOURNAL ON COMPUTER SCIENCE AND ENGINEERING, no 4, 2006

[9] Najmeh Ghanbari, mohamad Razavi, hasan Nabavi," A Smart Properties Selection Method Based on Binary Gravitational Search Algorithm in Persian Handwriting Number Recognition System", Electrical and Computer Engineering of Iran, 2011

[10] Kottakota Asish, P. Sarath Teja, R. Kishan Chander, Dr. D. Deva Hema, NeuroWrite: Predictive Handwritten Digit Classification using Deep Neural Networks, Artificial Intelligence,2023,https://doi.org/10.48550/arXiv.2311.01022

[11] Diaa s AbdElminaam email orcid 1; Farah Essam2; Hanein Samy2; Judy Wagdy2; steven Albert2, MLHandwritten Recognition: Handwritten Digit Recognition using Machine Learning Algorithms, Journal of Computing and Communication, Volume 2, Issue 1, January 2023, Page 9-19

[12] Danveer Rajpal & Akhil Ranjan Garg, Ensemble of deep learning and machine learning approach for classification of handwritten Hindi numerals, Journal of Engineering and Applied Science volume 70, 2023

[13] Nidhal Azawi, Handwritten digits recognition using transfer learning, Handwritten digits recognition using transfer learning, Volume 106,2023

[14] CvejicE, Prosody off the top of the head: Prosodic contrasts can be discriminated by head motion, Computers & Electrical Engineering,2024

[15] ZhangS, Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition Knowl-Based Syst, Computers & Electrical Engineering,2022

[16] LeeC.jun cowich, Emotion recognition using a hierarchical binary decision tree approach, Speech Commun ,2023

[17] X. Lei, H. Pan and X. Huang, "A Dilated CNN Model for Image Classification", IEEE Access, vol. 7, pp. 124087-124095, July 2022

[18] X. Jiang, Y. Wang, W. Liu, S. Li and J. Liu, "CapsNet CNN FCN: Comparative Performance Evaluation for Image Classification", International Journal of Machine Learning and Computing, vol. 9, no. 6, December 20121

[19] esmaiel miri, mohamad razavi, javad razavi, "The effect of clustering on the recognition of Persian manuscript cultivars with fuzzy classifier" ,2016

[20] ehsan jabir, reza jabir, reza ebrahimpour," A combination of two-class clauses for recognizing Persian manuscript cultivars", 16th Iranian Electrical Engineering Conference, 2007

[21] Corinna Cortes, Mehryar Mohri, Umar Syed, Deep Boosting, e 31 st International Conference on Machine Learning, Beijing, China, 2014

[22] Jafar tanha, Ensemble approaches to semi-supervised learning, UvA-DARE (Digital Academic Repository),2013, http://hdl.handle.net/11245/1.393046

[23] Ji Zhu, Hui Zou, Saharon Rosset and Trevor Hastie, Multi-class AdaBoost, Statistics and Its Interface Volume 2, 2009 349–360

[24] Freund, Y. and Schapire, R. A decision theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 1997 119–139 MR1473055

[25] Friedman, J. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 2001 1189–1232. MR1873328

[26] Friedman, J., Hastie, T., and Tibshirani, R. Additive logistic regression: a statistical view of boosting. Annals of Statistics 2000, 337–407. MR1790002

[27] Jamshid bagherzadeh, hasan asil, proposing a New Method of Image Classification Based on the AdaBoost Deep Belief Network Hybrid Method, TELKOMNIKA, Vol 17, No 5, 2019

[28] N-ary Decomposition for Multi-class Classification. Machine Learning Journal (MLJ), 2019

[29] Learning with Annotation of Various Degrees, IEEE Transactions on Neural Network and Learning Systems (TNNLS), 2019.

[30] Farrokhi, Alireza and Razavi, Seyed Nasser, Recognition of handwritten digits using deep learning, International Conference on Nonlinear Systems and Optimization of Electrical and Computer Engineering, 2014, https://civilica.com/doc/383305

[31] Mehdi Noroozi and Paolo Favaro. "Unsupervised learning of visual representations by solving jigsaw puzzles". In European Conference on Computer vision. Pages 69–84. Springer. 2016

[32] Karen Simonyi and Andrew Zisserman. Very deep convolutional networks for large scale image recognition. ArXiv preprint arXiv: 1409.1556, 2014.

[33] Pourqasem, Hossein and Hel Forosh, Mohammad Sadegh and Daneshvar, Sablan, Semantic classification of text images based on text value model, 2017, https://civilica.com/doc/1372237

[34] Aslimi Zamanjani, Javad and Shakur, Mohammad Hossein and Rahmani, Mohsen, Classification of noisy texture images using deep neural network and complete local binary pattern, 2021, https://civilica.com/doc/1362855

[35] Shahabinejad, Athara and Iftikhari, Mehdi, Image classification using deep convolutional neural networks based on distributed attention and Bayes inference, 5th National Technology Conference in Electrical and Computer Engineering, 2021, https:// civilica.com/doc/1281540

[36] Babaian, Vahidah and Madiri, Shaghaigh and Behlgardi, Seyedah Kausar, Classification of images using deep neural networks, The 5th National Conference on the Application of New Technologies in Engineering Sciences, Torbat Heydarieh, 2018, https://civilica.com/doc /1202833

[37] Transfer Hashing: From Shallow to Deep, IEEE Transactions on Neural Network and Learning Systems (TNNLS), 2018

[38] Harri Valpola. From neural PCA to deep unsupervised learning. In Adv. in Independent Component Analysis and Learning Machines, pages 143–171. Elsevier, 2015.arXiv:1411.7783.

[39] Learning Common and Feature-Specific Patterns: A Novel Multiple-Sparse-Representation-Based Tracker. IEEE Trans. Image Processing 27(4): 2022-2037 (2018)

[40] Aaron vandenOord, NalKalchbrenner, Pixel Recurrent Neural Networks, international Conference on Machine Learning, New York, NY, USA, 2016 ,2016, arXiv:1601.06759v3

[41] Jamishid Bagherzadeh, Hasan Asil, A review of various semi-supervised learning models with a deep learning and memory approach, Iran Journal of Computer Science, 2018

[42] Alex Krizhevsky, IlyaSutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. IN Advances in neural information processing systems, pages 1097–1105, 2012.

[43] Christian Szegedy, Wei Liu, YangqingJia, Pierre Sermanet, Scott Reed, DragomirAnguelov, DumitruErhan,

[44] Andrej Karpathy, George Toderici, SankethShetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[45] Christian Szegedy, Wei Liu, YangqingJia, Pierre Sermanet, Scott Reed, DragomirAnguelov, DumitruErhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. ArXiv preprint arXiv: 1409.4842, 2014.

[46] M.M. Javidi Fatemeh Sharifizadeh Fatemeh Sharifizadeh," A Modified Decision Templates Method for Persian Handwritten Digit Recognition a Modified Decision Templates Method for Persian Handwritten Digit Recognition", Journal of American Science, 2012