In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 11, No.1, January-March 2023, Serial Number 41

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)

- Afsharirad, Majid, Kharazmi University,Tehran, Iran
- AleAhmad, Abolfazl, University of Tehran, Tehran , Iran
- Alijani, Ali, Islamic Azad University, Bandar Anzali, Iran
- Badie, Kambiz, Tehran University, Iran
- Broumandnia, Ali, Islamic Azad University, Tehran, Iran
- Farsi, Hassan, University of Birjand, South Khorasan, Iran
- Ghaffari, Ali, Islamic Azad University, Tabriz Branch, Iran
- Ghasemi, Majid, Islamic Azad University Shahr-e-Qods, Iran
- Gholami, Mohammad, Babol Noshirvani University of Technology, Mazandaran, Iran
- Hasan, Junayed, Universiteit van Ulsan, Ulsan, South Korea
- Hoseini Shakib, Mehrdad, Islamic Azad University Karaj Branch, Iran
- Kalantarnia, Ali, ACECR, Tehran, Iran
- Kasaei, Shohreh, Sharif University, Tehran, Iran
- Mahdieh, Omid, University of Zanjan, Zanjan, Iran
- Minoofam, Seyed Amir Hadi, Qazvin Islamic Azad University, Qazvin, Iran
- Mirzaei, Abbas, Islamic Azad University, Ardabil, Iran
- Moayedi, Fatemeh, University of Larestan Higher Education Complex, Fars, Iran
- Moradi, Gholamreza, Amirkabir University,Tehran, Iran
- Mohammadi, Mohammad Reza, Iran University of Science and Technology, Tehran, Iran
- Mohammadi, Mukhtar, Lebanese French University, Kurdistan Region, Iraq
- Omid Mahdi, Ebadati, Kharazmi University, Tehran, Iran
- Pashazadeh, Saeid, Tabriz University, Tabriz, Iran
- Qingliang, Zhao, Northwestern University, Evanston, United State
- Rahman, Saifur, Virginia Tech,Riva San Vitale, Switzerland,
- Rasi, Habib, Shiraz University of Technology, Shiraz, Iran
- Saadatfar, Hamid, University of Birjand, Iran
- Samsami Khodadad, Farid, Amol University of Special Technology, Amol, Iran
- Shirmarz, Alireza, Al Taha University, Tehran, Iran
- Tanhaei, Mohammad, Ilam University, Ilam , Iran
- Teymoori, Mehdi, Zanjan University, Iran
- Tourani, Mahdi, University of Birjand, South Khorasan, Iran
- Yaghoobi, Kaebeh, Ale Taha Institute of Higher Education, Tehran, Iran

Table of Contents

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 1-11
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Breast Cancer Classification Approaches - A Comparative Analysis

Mohan Kumar[1*], Sunil Kumar Khatri[2], Masoud Mohammadian[3]

[1]. Amity Institute of Information Technology, Amity University, Noida (UP), India
[2]. Director of Campus, Amity University Tashkent, Tashkent City, Uzbekistan
[3]. Faculty of Science and Technology, University of Canberra, Canberra ACT Australia

Received: 12 Jan 2022/ Revised: 30 Mar 2022/ Accepted: 26 Apr 2022

Abstract

Cancer of the breast is a difficult disease to treat since it weakens the patient's immune system. Particular interest has lately been shown in the identification of particular immune signals for a variety of malignancies in this regard. In recent years, several methods for predicting cancer based on proteomic datasets and peptides have been published. The cells turns into cancerous cells because of various reasons and get spread very quickly while detrimental to normal cells. In this regard, identifying specific immunity signs for a range of cancers has recently gained a lot of interest. Accurately categorizing and compartmentalizing the breast cancer subtype is a vital job. Computerized systems built on artificial intelligence can substantially save time and reduce inaccuracy. Several strategies for predicting cancer utilizing proteomic datasets and peptides have been reported in the literature in recent years.It is critical to classify and categorize breast cancer treatments correctly. It's possible to save time while simultaneously minimizing the likelihood of mistakes using machine learning and artificial intelligence approaches. Using the Wisconsin Breast Cancer Diagnostic dataset, this study evaluates the performance of various classification methods, including SVC, ETC, KNN, LR, and RF (random forest). Breast cancer can be detected and diagnosed using a variety of measurements of data (which are discussed in detail in the article) (WBCD). The goal is to determine how well each algorithm performs in terms of precision, recall, and accuracy. The variation of each classification threshold has been tested on various algorithms and SVM turned out to be very promising.

Keywords: Artificial Intelligence; Machine Learning; Wisconsin Breast Cancer Diagnostic (WBCD) Dataset; k-Nearest Neighbors (k-NN): Support Vector Classifier: Logistic Regression; ExtraTree-Decision; Random-Forest

1- Introduction

One in eight women will be diagnosed with breast cancer throughout their lifetime, making it the leading cause of mortality in women[1]. Machine learning algorithms for early breast cancer diagnosis and prognosis have evolved as a result of improved technology and more access to data. Breast cancer is the largest cause of death for women, according to a WHO report, and the number of deaths is increasing every year [2]. Males are more prone than females to succumb to cardiac disease [3]. Breast cancer generally affects women over 40, but some risk factors make it possible for it to attack younger people as well [4]. It is impossible to make rational conclusions with traditional medical techniques and cannot handle enormous amounts of data without computational tools.

There are numerous systems that gather, analyse, and learn from cancer data, which explains why cancer data sources are so diverse. Even for data analysis, there are open-source tools that can be employed. The sheer volume of data involved makes data mining techniques very useful for early cancer detection. The earlier cancer is detected, the more accurate the diagnosis will be. For patients, improved therapy offers a higher probability of survival. Using previous data, classification models are built and trained for a given set of applications. This learning model reduces the amount of time and effort needed to manually predict and classify fresh data.

It also helps to detect difficult-to-perceive patterns from large and noisy datasets using machine learning algorithms by extracting knowledge and expertise from previous experiences. Experts believe that abnormal cell growth in the breast tissue is the cause of breast cancer in patients.

✉ **Mohan Kumar**
mohan_srivastava@hotmail.com

[5]. Figure 1 depicts a machine learning-based classification model for breast cancer. Before the characteristics can be extracted and the classification model constructed, the breast image database must be loaded. Benign non-cancerous conditions are those that are not life-threatening. Malignant cancer begins with abnormal cell proliferation and can spread quickly or even permeate neighboring tissues, making it somewhat lethal.[6].

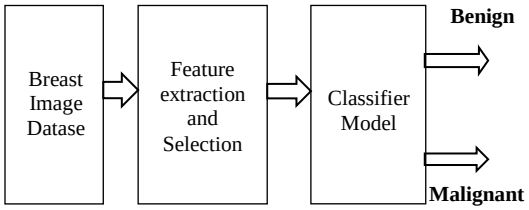


Fig. 1 Breast cancer image classification model [6]

It is hoped that the results of this study would help in the diagnosis of breast cancer by providing a classification model for malignant and benign breast cancers. Classification accuracy, Sensitivity, Specificity, and Precision are all assessed on the Wisconsin Breast Cancer Dataset (WBCD) [7] using k-nearest neighbors (k-NN), Logistic Regression, Support Vector Classifier (SVC), ExtraTree Decision (EDT), and Random Forest (RF). The most accurate models are LR, SVC, and DT, according to the results.. Seventy percent of the data is used for training, while the other thirty percent is used for testing. Data is divided in a 7:3 (70:30) ratio into training and testing portions because cross-validation is not a possibility here[8-9], the algorithms for classifying data performed admirably on the classification task.

The assessment of independent data set is analyses statistically using Cross validation technique for assessing. The evaluation of machine learning models is performed by model training on subsets of data input and available input data. Cross validation has a utility in scenario where there is a need of over-fitting. Optimal parameters are defined by cross validation. In current work cross validation not proved perfect for predicting over fitting in model. It was forcing model to get train again and again from scratch so current research work avoided the use of cross validation.

2- Literature Review

On this page, you'll learn about medically-approved machine learning algorithms used in breast cancer detection. Mangasarian and others [10] have used clinical data points to foresee the presence of cancer cells by forecasting the time interval at which recurrence will occur. Few of the studies conducted so far [11–13] also put forth work related to prediction of cancer and it's diagnosis by use of ML techniques such as decision tree to predict and diagnose cancer. According to Jin [14], KNN (K-Nearest Neighbor) is a classification algorithm used often among all machine learning approaches due to its simplicity and flexibility during implementation. The Wisconsin Prognostic Breast Cancer (WPBC) dataset was used by Belciug and colleagues [15] to compare K-means, cluster networks, and Self-Organizing Maps for the identification of breast cancer using the Wisconsin Prognostic Breast Cancer (WPBC) dataset [16], and the results showed that k-means were superior.

The Cancer Dataset was utilized by Chaurasia and Pal [17] prediction of breast cancer recurrence using ANNs, or artificial neural networks, Dyadic decision trees (DDTs), and Logistic Regression (LR) is implemented to retrieve accurate prediction about breast cancer. Research work outcome is the most accurate classifier for predicting the primary site of cancer development, Angeline [18] used the same dataset as Wisconsin Breast Cancer (WBC) and tested the performance of various algorithms such Nave Bayes, Decision tree (C4.5), KNN, and Support Vector Machine. When compared to the other approaches, Support Vector Machine performed the best. Author proposed and considered SVM as best classifier because it performs with high performance based on generalization as it doesn't add any prior knowledge. The function performed very well in high dimensional space of input dataset. This is the reason why SVM is being most promising as best classification function and has easily differentiate between instances of the two classes of breast cancer from training data.

Fuzzy clustering techniques were utilized by Abonyi and Szeifert [19] to predict cancer from the same dataset. the Wisconsin Diagnostic Breast Cancer Dataset (WDBC). Lavanya and coworkers [20] employed a dynamic, hybrid technique to improve classification accuracy, and the dataset was cross-validated 10 times. The hybrid technique included various function in combination and results turned out very accurate and model performed well. Results from [21] research show that using a validation set to predict a model's performance in Machine Learning, cross-validation is the best technique. The cross-validation when applied to dataset proved very powerful tool. The data set is best utilized and shown comparably very good performance. There are some very complex machine learning models, cross validation help in using same data in each step of the pipeline.

For his studies, Cruz [22] investigated cancer detection and prognosis machine learning strategies. Using a microarray of cancer data, Tan A.C. [23] examined the bagged decision tree, also known as the C4.5 decision tree [24], he used bagging methods, decision trees, neural networks, and the Support Vector Machine to analyze a cancer database. According to the report's findings, bagging approaches have higher levels of accuracy. As evidenced

by the publications listed above, breast cancer can be diagnosed using a number of machine learning (ML) approaches. Comparing these approaches to discover the most efficient algorithm is difficult due to the use of various datasets and data sizes. The algorithms can only be employed on a single platform and foundation if they are implemented on a single dataset and with a single data size. This data will be used to develop the most accurate algorithm for classifying the presence of breast cancer. Our objective is to determine the most effective machine learning algorithm for the provided dataset.

3- Literature Review

Mammary cancer is classed as benign or malignant based on the results of tests such as the Komen Breast Cancer Early Detection and Management System (MBC-ELDS). Breast cancer is more accurately and efficiently classified as benign or malignant by employing the methods outlined above.

3-1- Proposed System Model

WBC: In this study, the dataset used was Wisconsin Breast Cancer, which can be obtained at the UCI Machine Learning Repository ((WBC). The pre-processed data will be fed into a variety of machine learning algorithms, including SVM, ExtraTree Decision, KNN, Random Forest, and Logistic Regression, to predict the likelihood of a cancer diagnosis. Measurement of performance and accuracy is the goal.

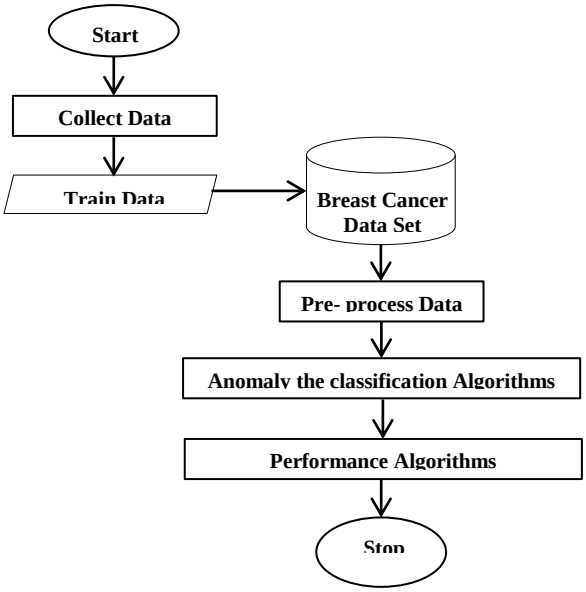


Fig. 2 Steps in proposed Methodology [25].

to do this, Machine Learning approaches make use of performance indicators such as accuracy, precision, sensitivity, and specificity. There will be a suitable categorization model proposed and validated. The proposed system model necessitates the completion of various steps. Fig.2 describes the method which has been conducted in the analysis.

3-2- Dataset

Diagnostic (WBCD) dataset, which was found in the UCI machine learning repository [7],[26]. For three years, researchers from the University of Wisconsin Hospitals (Wolberg, Street, and Mangasarian) collected this data.

We're utilizing the binary classification dataset for this exercise (WBCD). Within this collection, there are 569 records with 32 unique attributes. Breast lumps are aspirated using a fine needle aspirate (FNA) and then images of those aspirates are scanned for use in this dataset. The characteristics are as follows: Ten separate digitized cell nuclei features, mean and standard error, and the worst outcome for those ten diagnoses are included in these 30 real-value input feature attributes (Real, Positive). Type (B = benign, M = malignant) of cancer. Radius, Smoothness, Texture, Compactness, Fractal Dimension, Concavity, Perimeter, Convex Points, Symmetry, and Area can all be listed as qualities. [27]. The diagnoses were included in the dataset, which covered a wide spectrum of lumps and masses. A diagnosis is given when a tumor or lump is determined to be malignant or benign (B). There are four significant digits in these property values. Figure 3 reveals that 357 of the sample values are regarded as benign, whereas 212 are regarded as malignant.

Table 1: Wisconsin Dataset

Attributes	***Number of Attributes***
Sample Total	569
Dimensionality	30
Classes	2
Sample per class	Benign: 357 (62.74%) Malignant: 212 (37.26%)
Feature	Real Positive

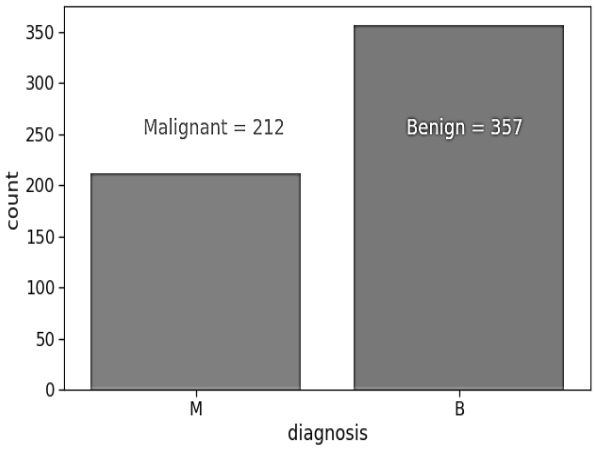


Fig. 3 Class distribution of WBCD data.

Heat maps in Figure 4 show the relationships between the variables. The correlation between two datasets shows the relationship between their attributes changing. A positive correlation occurs when two features move or change at the same time. If they go in the opposite direction, they are said to be negatively related. It reveals which traits are closely related to one another. It's critical to understand this since some machine learning algorithms do not work as expected.

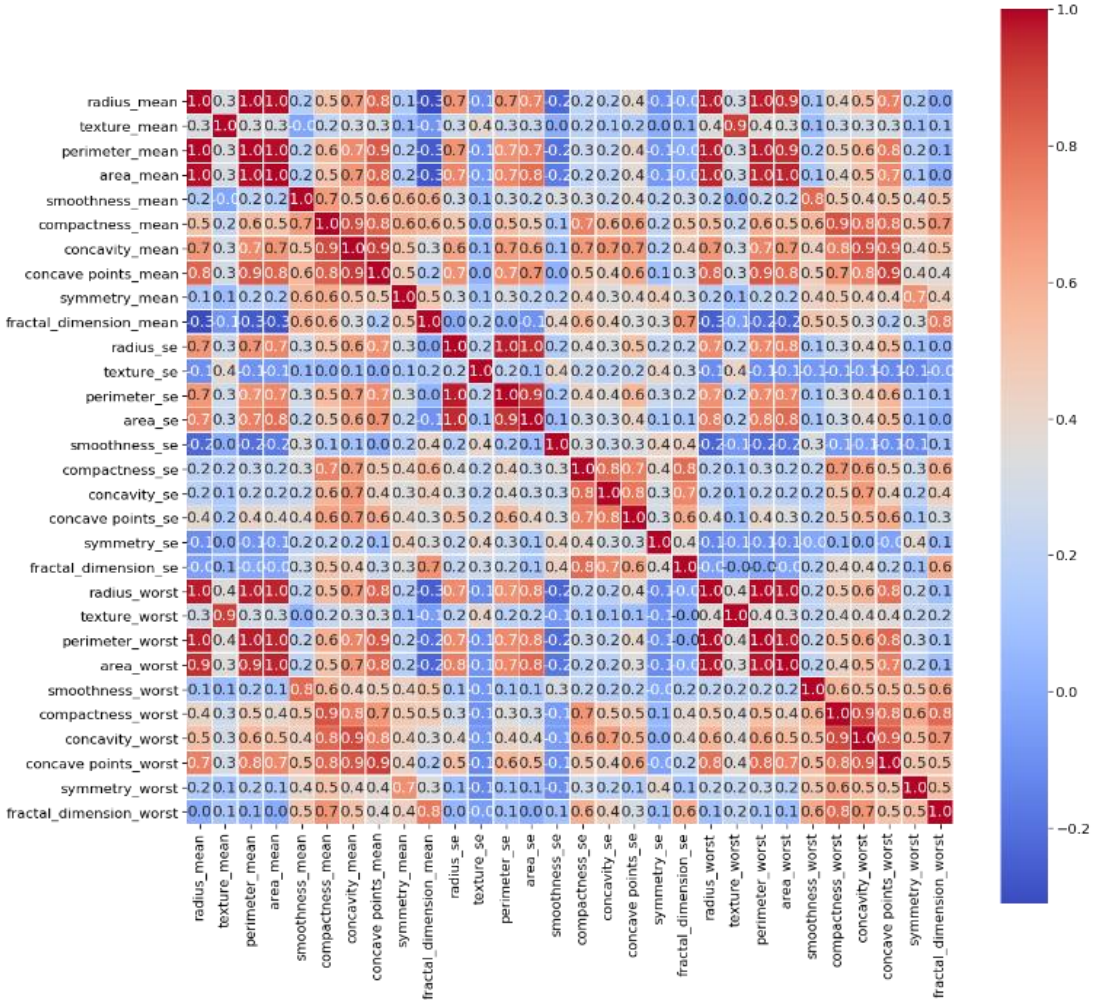


Fig. 4 Correlations between the features.

Many input features are highly linked in their data. The confusion matrix in Fig 4 showing collinear variables. Multi collinear variables are removed by few identified highly correlated variables which are independent. The combination of linear independent variables are summed up for performing analysis on variables which are highly correlated variables by applying principal components analysis or partial least squares regression.

These two ideas can be combined in a zillion different ways. Frequently, high-connectivity features are paired with ones that provide inaccurate information. By removing the most relevant features, we can get rid of the discriminating information they hold. We also see from this that while pushing birth control or other therapies we must remember that just because they can predict the outcome does not mean that they are at their core is an issue.

3-3- Preprocessing

Transformations are employed in pre-processing before the data is fed into the algorithm. the procedure used to collect the data Pre-processing is employed in order to turn unstructured data into something that can be read by machines. Standard methods for machine learning make the assumption that more data equates to better outcomes. In order to be used in this study, the WBCD data had to be normalized.

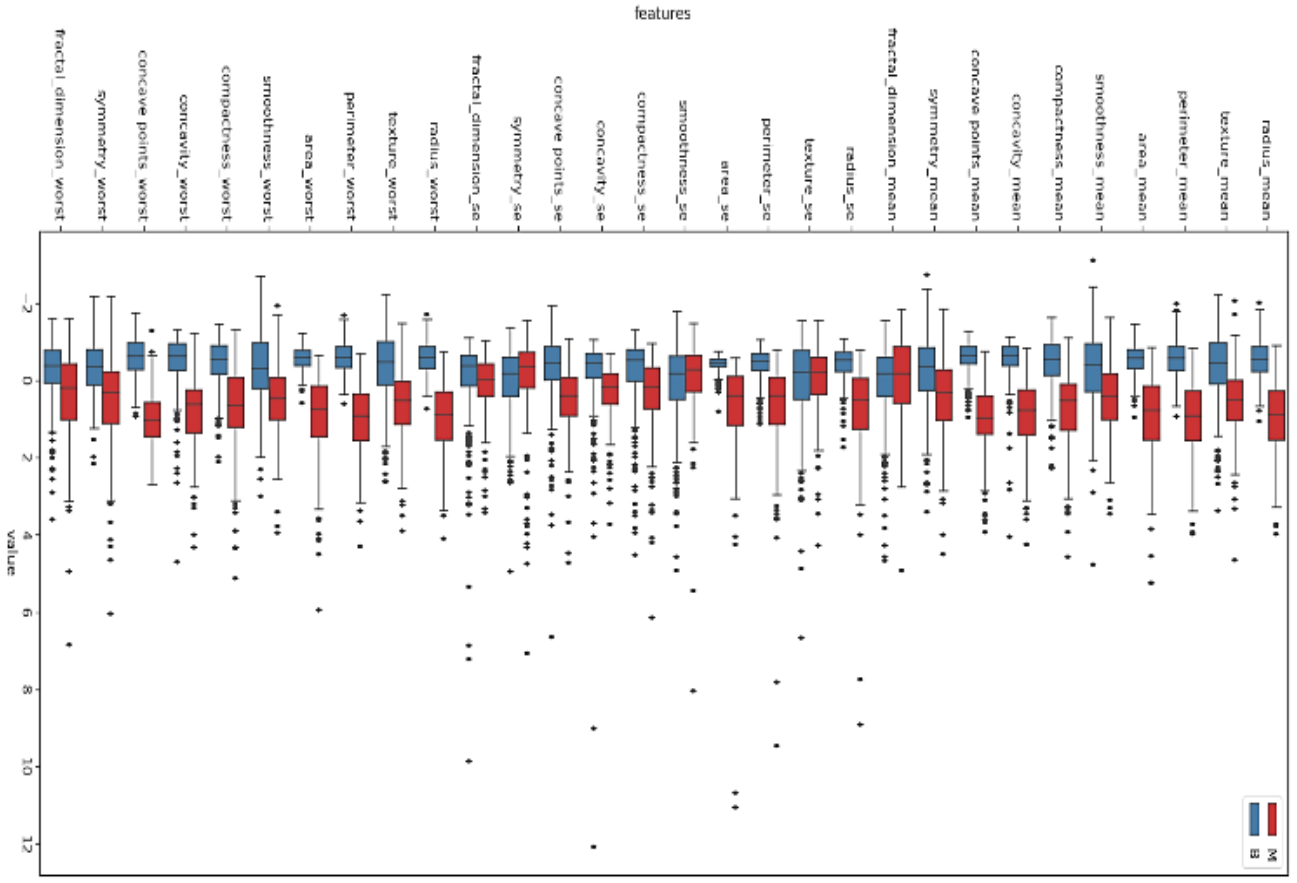


Fig. 5 Boxplot of features by diagnosis.

To produce the standard Gaussian distribution, several means with varying standard deviations are employed to turn the characteristics into a standard Gaussian distribution with a mean of 0 and a standard deviation of 1. A distinct ID is assigned to each patient in the database to represent them in the dataset. The first attribute id, does not include any references to this variable. The next consideration is the diagnostic value. The recommended algorithms should be able to predict the value of this variable. A Z score denotes a feature's standardized value. How to compute Z-score: follow these steps.

$$Z = (x - \mu)/(\sigma) \quad (1)$$

x denotes the standardized value, whereas,, and the stddev are used to express the mean and standard deviation, respectively. After standardizing all numerical values, we create a box plot (Fig. 5) to organize the data. Because there were so many outliers in the data, numerous features had to be discarded. As a result, when analyzing and categorizing the boxplot, we failed to take advantage of these features. There are some data points in analysis which are very far from other data points.

The outlier retrieved is very much different from other observations on axis. There are measurements which are very much different. There are experimental errors in the form of outliers. The analysis become very difficult when we try to apply statistic only because of outliers. So sometimes it is much needed to avoid them.

This chart shows how similar the graphics are amongst several of the features. There appears to be some similarity when comparing the perimeter with the area or the perimeter with the area se. However, certain traits differ based on whether a tumor is diagnosed as malignant or benign. Examples of the same include the concavity mean, radius means, and area means. However, if we look at metrics like texture se or fractal dimension mean, we can detect similarities in the distribution of malignant and benign tumors.

4- Nonlinear Machine Learning Algorithms

There are two types of machine learning algorithms: supervised and unsupervised. When using supervised learning, we provide a dataset to the machine to train it, and then the program outputs the accurate answers. Unsupervised learning, on the other hand, does not have any predetermined data sets or goals [28]. Machine learning has gained remarkable practical success in a variety of different fields. Medical expert systems [29], [30], social networks [31], and speech recognition [32], [33] are a few of the most common applications. As a result of past observations, machine learning learns how to make exact predictions. Classification is a supervised learning-related topic of study. The goal, as the name suggests, is to classify the samples in order to come up with a collection of plausible classes that are discrete. These include k-nearest neighbors (k-NN), support vector classifier (SVC), logistic regression, and extra-tree decision) and random-forest (RF) [34].

Classification involves educating the classifier to be aware of the different classes. As an added bonus, categorization is basically a mission to anticipate the value of a specific discrete characteristic. When it comes to classification, supervised learning algorithms are the go-to solution. This data is collected for the computer to learn a classification model from. After that, future projections will be possible. The classification problem is best solved with an algorithm based on supervised learning. Five different classification algorithms were used to sort the Wisconsin dataset [35]. To sum up, here are the five methods

4-1- K-Nearest Neighbors (K-NN)

It's possible to make predictions with supervised learning, but not with unsupervised learning. Depending on your requirements, it can be used for regression or clustering. The number of nearest neighbors is denoted numerically by a letter like K. There doesn't appear to be any formal training programme in place, as far as I'm aware. As part of the forecasting process, the Euclidean distance between the k-nearest neighbors is first taken into account [36].. When it comes to breast cancer, we've already labelled the cases as either malignant or benign. Labels are classified according to how far apart they are from one another. Figure 6 explains the KNN algorithm's technique better than any other figure.

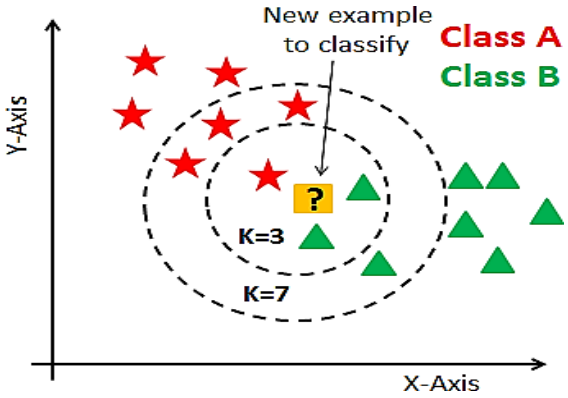


Fig. 6 Representation of KNN algorithm [26].

4-2- Logistic Regression (LR)

Logistic regression was first used in biology study in the early twentieth century. Using it in social studies classes is

becoming more and more common. This is also covered by predictive analysis. When there is only one dependent variable, it is most often used. To distinguish between a linear and a logistic regression, we employ the dependent variable. For continuous variables, the most common approach is linear regression, as shown in Figure 7. The following section outlines the many steps involved:

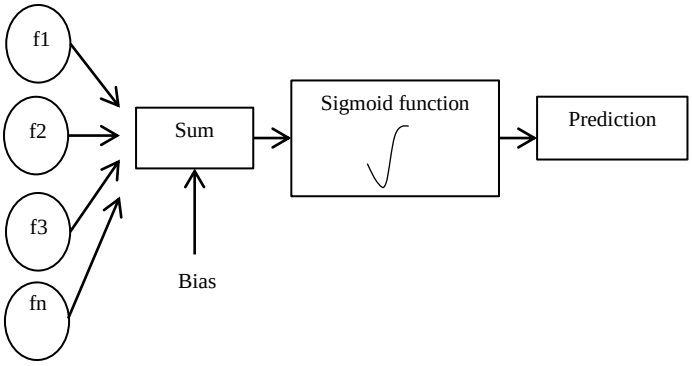


Fig. 7 Visualization of logistic regression steps.

Logistic regression makes use of both forward and backward propagation. Weight and characteristics are multiplied to begin the forward propagation phase. There's no way to tell how much something weighs when it's in its infancy. Because of this, weights can be assigned to certain random values. For sigmoid functions, the probability range is 0 to 1. According to the set threshold value, the forecasting method is carried out. A comparison of the projected and actual values is performed following the completion of the forecasting process. It's possible to calculate the difference between the projected and actual values using a loss function. The backward propagation approach is utilized if the loss function value is extremely high. By using backward propagation [38], one can obtain a derivative and update the weight values in accordance with the cost function. The sigmoid function is depicted in the following diagram:

$$\text{sigmoid}(x) = \frac{e^x}{1 + e^x} \quad (2)$$

4-3- Extra Tree-Decision (ETD)

Classifiers like the Extra Tree classifier are useful for combining the results of different decision trees that aren't related in any way. Foresters collect and sort these twigs to create different kinds of trees in the area. All of the decision trees have benefited from the use of genuine training data. If the loss function value is too large, the backward propagation strategy is used. A derivative is obtained and weight values depending on the cost function are updated using backward propagation [38]. Because of this, every call tree must start with the most basic features, which are then segregated from the data using statistical criteria known as the Gini Index. This random selection of options leads to a number of decision trees with no correlation. [39].

4-4- Support Vector Classifier (SVC)

SVMs, or Support Vector Machines, are frequently employed in machine learning. An N-dimensional hyper plane was sought for by these algorithms in order to classify the incoming data. Many hours are devoted to devising a plan that will maximise profits. When employing feature numbers, it is possible to distribute the N dimensionally uniformly. You can easily compare the two features. Despite the abundance of tools available, categorising things isn't always straightforward. To get the most accurate forecast, we should try to widen the margin of error. [40] This may be seen in Figure 8 where it appears that the SVM is:

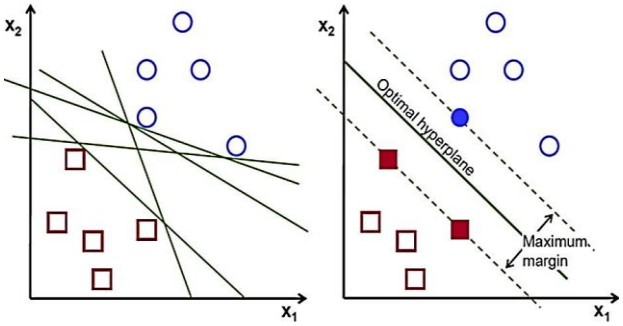


Fig.8 Support vector visualization [41].

Being over-classified often means being under-classified, and this must always be remembered. The error margin could be exceedingly tiny if we choose a categorization algorithm that does not sacrifice any of the individual samples. Final results may be less precise as a result of this situation. Support vectors that are most closely related to the hyper plane can also be considered as long as the gap between classes is as big as possible.

4-5- Random Forest (RF)

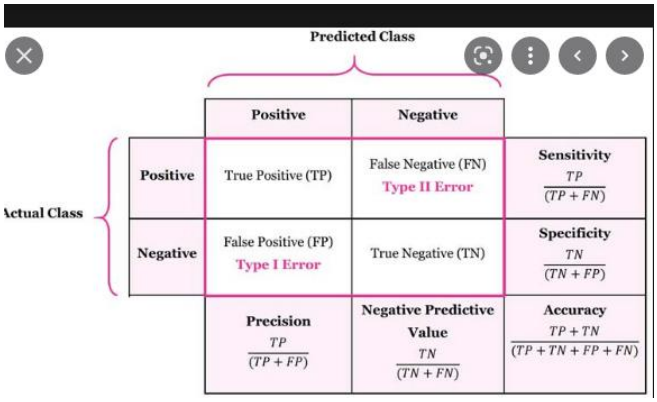
After the random forest, we come to an ensemble-learning model called the Random forest. This data analysis technique can be used for both regression and classification. It's termed a random forest since it's made up of many different decision trees. That is why utilizing a random forest instead of the more traditional decision tree seems sensible in some circumstances. It is necessary to create a classifier that uses the process of extracting these features[42], given that the rotating forest technique is being applied. After that, this attribute collection is divided up into K different groups using a randomization algorithm. Classifiers gain significance and precision in this way. [43].

The decision tree's most difficult element is selecting characteristics, and there are several ways to go about it.

Instead than seeking the most essential feature when dividing nodes, random forest seeks for the best features within a completely random collection of attributes. Even if it's curated even more haphazardly in the future, it's still doable. So it's plausible since, while using Random Forest, random attributes might be sought out rather than the most important traits for node splitting among the better ones. [44]. An internal function parameter can increase model accuracy or speed. The accuracy of the prediction model can be improved by utilizing characteristics such as Max features, minimum sample leaves, and n estimators, among others. When running the models, we frequently use N jobs and a randomized state machine in order to speed them up (RSM). Researchers employed characteristics including n estimators and random state parameters to calculate how many trees should be planted to improve the model's accuracy and speed.

5- Proposed System Model

Evaluation of the Model: Several performance measures can be used to assess the accuracy, sensitivity, specificity, and precision[45] of each Machine Learning approach. There are four ways to look at these choices: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) [46]. The true positive decision happens when the instances of malignancy are predicted aptly. True negative decisions are an occurrence that occurs when instances of benign behavior are correctly predicted. In cases where a benign disease is predicted to be malignant, the False Positive decision is made [47]. A false negative occurs when a malignant disease is predicted to be benign [46].



i. Accuracy can be calculated as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

ii. Sensitivity can be calculated as:

$$\text{Sensitivity (recall)} = \frac{TP}{TP+FN}$$

iii. Specificity can be calculated as:

$$\text{Specificity} = \frac{TN}{TN+FP}$$

iv. Precision can be calculated as:

$$\text{Precision} = \frac{TP}{TP+FP}$$

After these variables (Accuracy, Sensitivity, Specificity, and Precision) have been determined, a confusion matrix is created utilizing these values. When there is rise in classification threshold it will decrease the false positives, which further increased the precision. When there is decrease in raising the threshold of classification reduce the false positives which will further increase precision.

6- Result and Discussion

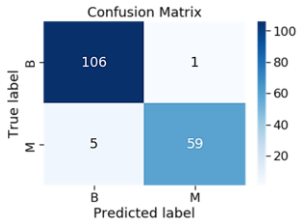
About 70% of the dataset items are used for training, while the remaining 30% are omitted for testing. Using a random number generator, these two subgroups are created from the entire dataset. Four separate measures are used to evaluate the performance of machine learning algorithms: specificity, sensitivity, precision, and accuracy (or accuracies). To determine the algorithm's overall performance, various measurements are compared. Multiple machine learning algorithms used to classify breast cancer data as benign or malignant, and then present the results in a confusion matrix (see Table II). Table III summarizes the various machine learning approaches to the breast cancer dataset based on various performance criteria. Table III. k-NN, support vector classifier, logistic regression, additional tree decision and random forest classification outcomes were evaluated. This highlights the areas of interest for various classification techniques. [48].

Tables III show that the SVC and ExtraTree classifiers are the most precise, specific, and accurate. On the other hand, radio frequency (RF) technology boasts the highest sensitivity. Results show that SVM and ExtraTree models identify tumors as benign or malignant with the greatest accuracy.

Table 2: Confusion Matrix.

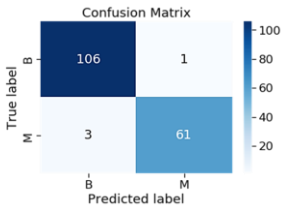
(i) KNN

KNN	Test Outcome	
Predictions	Benign	Malignant
Benign	106(TP)	1(FN)
Malignant	5 (FP)	59 (TN)



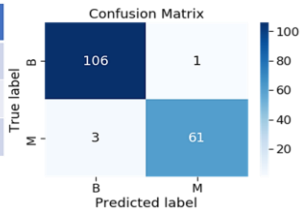
(ii) SVM

SVM	Test Outcome	
Predictions	Benign	Malignant
Benign	106(TP)	1(FN)
Malignant	3 (FP)	61 (TN)



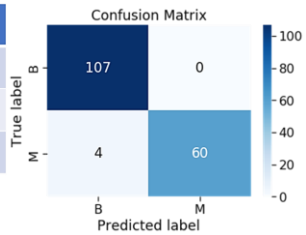
(iii) ExtraTree

ExtraTree	Test Outcome	
Predictions	Benign	Malignant
Benign	106(TP)	1(FN)
Malignant	3 (FP)	61 (TN)



(iv) Random Forest

Random Forest	Test Outcome	
Predictions	Benign	Malignant
Benign	107(TP)	0(FN)
Malignant	4 (FP)	60 (TN)



(v) Logistic Regression

Logistic Regression	Test Outcome	
Predictions	Benign	Malignant
Benign	102(TP)	5(FN)
Malignant	3 (FP)	61 (TN)

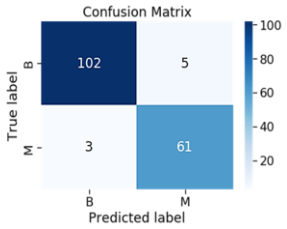


Table 3: Algorithm's comparison

	KNN		SVM	ExtraTree	RF	LR
Accuracy (%)	96.4		**97.7**	**97.7**	**97.7**	95.3
Sensitivity (%)	99.0		99.0	99.0	**100**	95.3
Specificity (%)	92.1		**95.3**	**95.3**	93.7	95.3
Precision (%)	95.4		**97.2**	**97.2**	96.3	97.1

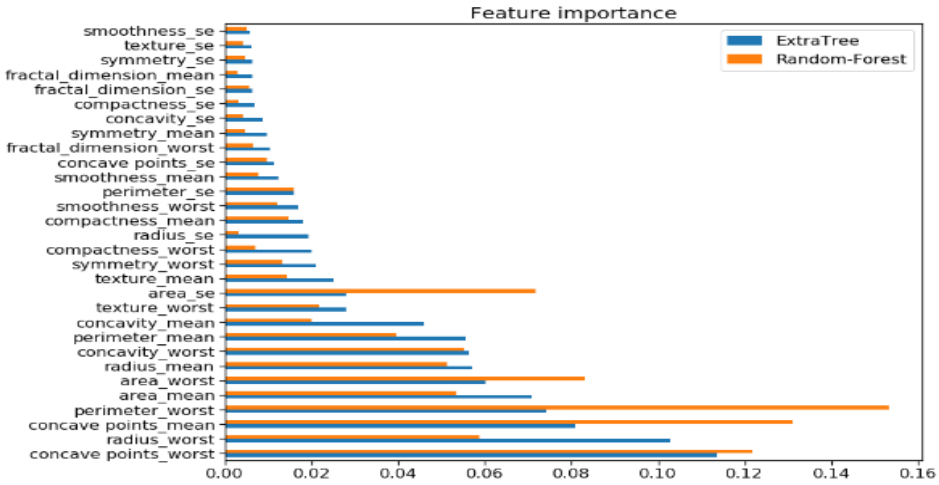


Fig.9 Feature Importance

Figure 9 represents feature selection which is most important for analyzing various machine learning algorithms

Table 4: Important Features

perimeter_worst	0.153370
concave points_mea	0.131221
concave points_worst	0.121852
area_worst	0.082954
area_se	0.071767
radius_worst	0.058753

The six most important features having data type float64 are mentioned on the above table

7- Conclusion and Future Scope

A number of machine learning algorithms, such as SVM, ExtraTree, Random Forest, and KNN, are put to the test. The analysis also makes use of Logistic Regression. Using a variety of machine learning methods, the goal was to discover the most accurate classifier. To sort the breast cancer dataset, all of the classifiers will be put to work (Wisconsin Breast Cancer Dataset). A Python tool was used to do the research. The accuracy, sensitivity, specificity, and precision of SVM and ExtraTree were superior to those of the other classifiers. Besides highlighting the six most essential qualities, the research also discusses how these characteristics affect the prediction models. We'll look into it some further. Alternative dataset combinations (such as 50-50, 80-20, 60-40, and so on) will be examined in a future study to see how well Bayesian Networks, Gradient Descents, and artificial neural networks perform when dimension reduction approaches are applied. The capabilities of KERAS-Neural-Networks Tensor Flow will be evaluated in the following publication by using deep learning that can handle enormous datasets, we'll see how deep learning compares to other approaches for handling large datasets. Because of this, knowing which method is best for which type of data will be much easier.

References

[1] D. Hanahan and R. A. Weinberg, “Hallmarks of Cancer: The Next Generation,” Cell, vol. 144, no. 5, pp. 646–674, Mar. 2011,

[2] S. Katuwal, P. Jousilahti, and E. Pukkala, “Causes of death among women with breast cancer: A follow- up study of 50 481 women with breast cancer in Finland,” Int. J. Cancer, vol. 149, no. 4, pp. 839–845, Aug. 2021

[3] S. F. Khorshid and A. M. Abdulazeez, “breast cancer diagnosis based on k-nearest neighbors: A review,” PalArch’s J. Archaeol. Egypt/Egyptology, vol. 18, no. 4, pp. 1927–1951, 2021

[4] Tawam Hospital | Medical News. (n.d.). Retrieved November 19, 2014, from http://www.tawamhospital.ae/english/news/print.aspx?NewsID= 367

[5] M. Karabatak, “A new classifier for breast cancer detection based on Naive Bayesian,” Measurement, vol. 72, pp. 32–36, 2015

[6] A-Al. Nahid and Y. Kong, “Involvement of machine learning for breast cancer image classification: a survey,” Comput. Math. Methods Med., 2017

[7] M. Kumar, S. K. Khatri, and M. Mohammadian, “Breast cancer identification and prognosis with machine learning techniques-An elucidative review,” J. Interdiscip. Math., vol. 23, no. 2, pp. 503–521, 2020,

[8] A. Joshi and A. Mehta, “Comparative Analysis of Various Machine Learning Techniques for Diagnosis of Breast Cancer,” Int. J. Emerg. Technol., vol. 8, no. 1, pp. 522–526, 2017.

[9] B. Soni, A. Bora, A. Ghosh, and A. Reddy, “RFSVM: A Novel Classification Technique for Breast Cancer Diagnosis,” Int. J. Innov. Technol. Explor. Eng., 2019.

[10] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, “Breast cancer diagnosis and prognosis via linear programming,” Oper. Res., vol. 43, no. 4, pp. 570–577, 1995

[11] Zhi-H. Zhou and Y. Jiang, “Medical diagnosis with C4. 5 rule preceded by artificial neural network ensemble,” IEEE Trans. Inf. Technol. Biomed., vol. 7, no. 1, pp. 37–42, 2003,

[12] T. Ornthammarath, “Artificial neural networks applied to the seismic design of deep tunnels,” Università degli Studi di Pavia, 2007.

[13] D. Delen, G. Walker, and A. Kadam, “Predicting breast cancer survivability: a comparison of three data mining methods,” Artif. Intell. Med., vol. 34, no. 2, pp. 113–127, Jun. 2005

[14] Jini Marsilin, “An Efficient CBIR Approach for Diagnosing the Stages of Breast Cancer Using KNN Classifier,” Bonfring Int. J. Adv. Image Process., vol. 2, no. 1, pp. 01–05, Mar. 2012

[15] S. Belciug, A-B Salem, F. Gorunescu, and M. Gorunescu, “Clustering-based approach for detecting breast cancer recurrence,” in 2010 10th International Conference on Intelligent Systems Design and Applications, pp. 533–538, Nov. 2010,

[16] M.Lichman, “UC Irvine Machine Learning Repository,” 2015. http://archive.ics.uci.edu/ml.

[17] V. Chaurasia and S. Pal, “Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability,” Int. J. Comput. Sci. Mob. Comput., vol. 3, no. 1, pp. 10–22, 2014.

[18] Christobel, Angeline, and Y. Sivaprakasam, “An empirical comparison of data mining classification methods,” vol. 3, no. 2, 2011

[19] J. Abonyi and F. Szeifert, “Supervised fuzzy clustering for the identification of fuzzy classifiers,” Pattern Recognit. Lett., vol. 24, no. 14, pp. 2195–2207, 2003

[20] D. Lavanya, “Ensemble Decision Tree Classifier For Breast Cancer Data,” Int. J. Inf. Technol. Converg. Serv., vol. 2, no. 1, pp. 17–24, Feb. 2012

[21] Abad, Monica, James Carlisle Genavia, Jaybriel Lincon Somcio, and Larry Vea. "An Innovative Approach on Driver's Drowsiness Detection through Facial Expressions using Decision Tree Algorithms." In 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0571-0576. IEEE, 2021.

[22] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," Cancer Inform., vol. 2, p. 117693510600200030, 2006

[23] A. C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," 2003

[24] G. L. Tsirogiannis, D. Frossyniotis, J. Stoitsis, S. Golemati, A. Stafylopatis, and K. S. Nikita, "Classification of medical data with a robust multi-level combination scheme," in 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), 2004

[25] G. Sudhamathy, M. Thilagu, and G. Padmavathi, "Comparative analysis of R package classifiers using breast cancer dataset," Int J Eng Technol, vol. 8, pp. 2127–2136, 2016.

[26] Frank A. UCI machine learning repository. http://archive. ics. uci. edu/ml. 2010.

[27] WHO, WHO position paper on mammography screening. World Health Organization, 2014.

[28] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in 2016 5th international conference on electronic devices, systems and applications (ICEDSA), pp. 1–4, 2016

[29] I. Castiglioni et al., "AI applications to medical images: From machine learning to deep learning," Phys. Medica, vol. 83, pp. 9–24, Mar. 2021

[30] R. Abdlaty, J. Hayward, T. Farrell, and Q. Fang, "Skin erythema and pigmentation: a review of optical assessment techniques," Photodiagnosis Photodyn. Ther., vol. 33, p. 102127, Mar. 2021,

[31] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Heal. Inf. Sci. Syst., vol. 6, no. 1, p. 8, Dec. 2018

[32] M. D. Skowronski and J. G. Harris, "Acoustic detection and classification of microchiroptera using machine learning: Lessons learned from automatic speech recognition," J. Acoust. Soc. Am., vol. 119, no. 3, pp. 1817–1833, Mar. 2006

[33] L. Deng and X. Li, "Machine Learning Paradigms for Speech Recognition: An Overview," IEEE Trans. Audio. Speech. Lang. Processing, vol. 21, no. 5, pp. 1060–1089, May 2013

[34] "Cardio-Vascular Disease Prediction based on Ensemble Technique Enhanced using Extra Tree Classifier for Feature Selection," Int. J. Recent Technol. Eng., vol. 8, no. 3, pp. 3236–3242, Sep. 2019

[35] Kaggle.com, "Breast Cancer Wisconsin (Diagnostic) Data Set," 2021. https://www.kaggle.com/uciml/breast-cancer-wisconsin-data.

[36] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN Classification With Different Numbers of Nearest Neighbors," IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 5, pp. 1774–1785, May 2018

[37] M. Faisal, A. Scally, R. Howes, K. Beatson, D. Richardson, and M. A. Mohammed, "A comparison of logistic regression models with alternative machine learning methods to predict the risk of in-hospital mortality in emergency medical admissions via external validation," Health Informatics J., vol. 26, no. 1, pp. 34–44, Mar. 2020

[38] V. Tatan, "Your Beginner Guide to Basic Classification Models: Logistic Regression and SVM," 2019.

[39] O. Maier, M. Wilms, J. von der Gablentz, U. M. Kramer, T. F. Munte, and H. Handels, "Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences," J. Neurosci. Methods, vol. 240, pp. 89–100, 2015

[40] S. B. Kotsiantis, "Decision trees: a recent overview," Artif. Intell. Rev., vol. 39, no. 4, pp. 261–283, 2013

[41] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA. Cancer J. Clin., vol. 68, no. 6, pp. 394–424, Nov. 2018

[42] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," Expert Syst. Appl., vol. 134, pp. 93–101, Nov. 2019

[43] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 10, pp. 1619–1630, 2006

[44] M. Maria and C. Yassine, "Machine learning based approaches for modeling the output power of photovoltaic array in real outdoor conditions," Electronics, vol. 9, no. 2, p. 315, 2020

[45] S. S. Shajahaan, S. Shanthi, and V. ManoChitra, "Application of data mining techniques to model breast cancer data," Int. J. Emerg. Technol. Adv. Eng., vol. 3, no. 11, pp. 362–369, 2013.

[46] S. Raschka, "An overview of general performance metrics of binary classifier systems," arXiv Prepr. arXiv1410.5330, 2014

[47] Madooei, Ali, Ramy Mohammed Abdlaty, Lilian Doerwald-Munoz, Joseph Hayward, Mark S. Drew, Qiyin Fang, and Josiane Zerubia. "Hyperspectral image processing for detection and grading of skin erythema." In Medical Imaging 2017: Image Processing, vol. 10133, pp. 577-583. SPIE, 2017.

[48] R. Abdlaty et al., "Hyperspectral Imaging and Classification for Grading Skin Erythema," Front. Phys., vol. 6, Aug. 2018

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 12-23
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Developing A Contextual Combinational Approach for Predictive Analysis of Users Mobile Phone Trajectory Data in LBSNs

Fatemeh Ghanaati[1], Gholamhossein Ekbatanifard[2*], Kamrad Khoshhal Roudposhti[2]

[1]. Department of Computer Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran
[2]. Department of Computer Engineering, Lahijan branch, Islamic Azad University, Lahijan, Iran

Received: 09 Dec 2021/ Revised: 12 Feb 2022/ Accepted: 22 March 2022

Abstract

Today, smartphones, due to their ubiquity, have become indispensable in human daily life. Progress in the technology of mobile phones has recently resulted in the emergence of several popular services such as location-based social networks (LBSNs) and predicting the next Point of Interest (POI), which is an important task in these services. The gathered trajectory data in LBSNs include various contextual information such as geographical and temporal contextual information (GTCI) that play a crucial role in the next POI recommendations. Various methods, including collaborating filtering (CF) and recurrent neural networks, incorporated the contextual information of the user' trajectory data to predict the next POIs. CF methods do not consider the effect of sequential data on modeling, while the next POI prediction problem is inherently a time sequence problem. Although recurrent models have been proposed for sequential data modeling, they have limitations such as similarly considering the effect of contextual information. Nonetheless, they have a separate impact as well. In the current study, a geographical temporal contextual information-extended attention gated recurrent unit (GTCI-EAGRU) architecture was proposed to separately consider the influence of geographical and temporal contextual information on the next POI recommendations. In this research, the GRU model was developed using three separate attention gates to consider the contextual information of the user trajectory data in the recurrent layer GTCI-EAGRU architecture, including timestamp, geographical, and temporal contextual attention gates. Inspired by the assumption of the matrix factorization method in CF approaches, a ranked list of POI recommendations was provided for each user. Moreover, a comprehensive evaluation was conducted by utilizing large-scale real-world datasets based on three LBSNs, including Gowalla, Brightkite, and Foursquare. The results revealed that the performance of GTCI-EAGRU was higher than that of competitive baseline methods in terms of Acc@10, on average, by 42.11% in three datasets.

Keywords: LBSN; Trajectory Data; Contextual Information; GRU.

1- Introduction

Nowadays, people widely use location-based social networks (LBSNs) and enjoy location-based services (LBSs) using their mobile devices for sharing their locations with others by making check-ins at locations or points of interests (POIs) that they have visited, including shops, museums, and restaurants [1]. The massive record of users' check-in data provides a chance to conduct research on people's mobility behaviors, in particular, for POI recommendation systems [2,3]. In addition, governments can use predictions about people's future destinations and develop better transportation and scheduling strategies for alleviating traffic jams and handling crowd congestions [5,6,7,8]. Some geographical and temporal information exists in a user's historical check-in sequence [4,9], having different effects on recommending the next POI. In this study, it was attempted to separately consider this contextual information to better train the proposed model. Human mobility is extremely complex and diverse; therefore, many previous studies were unable to simply determine the offering of the next POI recommendation [4,6]. Matrix factorization (MF) and other collaborative filtering (CF) techniques have widespread use for recommending a list of personally ranked POIs to the users [2]. Typically, approaches to MF include contextual information about the user. This helps provide valuable recommendations to users who lack enough historical check-ins and is generally referred to as the cold-start problem. However, the employment of collaboration filtering (CF)-based methods complicates the processing of

✉ **Gholamhossein Ekbatanifard**
ekbatanifard@liau.ac.ir

sequence data and capturing of dynamic user's preferences [2,6,11]. As a result, the ongoing challenges lie in the manner of integrating the information of different features to accurately model users' complex behavioral preferences and then recommending reliable POIs [13].

Recurrent neural networks (RNNs) have recently been successfully applied to sequential recommender systems [1,4,8,15]. Thus, long-term dependencies can be captured by the hidden states of recurrent methods [4,16]. Many types of recurrent-based approaches have considered geographical and temporal factors to enhance the performance of POI recommendation algorithms [2,4,11,12,15]. Nonetheless, the present RNN-based POI recommendation methods face the alleviation of the cold-start problem [11]. In this regard, one of the excellent choices is to incorporate RNN-based POI recommendation methods with the MF method to enjoy the benefits of each one [2]. The user's historical check-in behaviors do not significantly pose any problems in predicting the next behavior; hence, it is necessary to take only the important information into serious consideration [1,11]. Therefore, the attention mechanism (AM) has been proposed to deal with this challenge. The AM can enhance the capability of the neural network in capturing long-term dependencies and boost the ability to interpret neural networks [18]. In this study, the idea of the AM was used to address the most important contextual information.

1-1- Motivations

This study focused on the next POI recommendation through modeling check-in sequences and considering geographical and temporal contextual influences separately and proposed a novel geographical temporal contextual information extended attention gated recurrent unit (GTCI-EAGRU) for the next POI recommendation. Among the recurrent models, the GRU model is highly simple and does not include many parameters in contrast to the long-short term memory (LSTM) model. In addition, this model can ignore the earlier unit hidden state, which is impossible with the traditional RNN [4,6]. Thus, a GRU network was developed to model check-in sequences while paying attention to geographical distances and time intervals between two successive check-ins [19]. It is noteworthy that any piece of contextual information needs individual consideration during modeling since the effects of contextual information on user behavior are different [2,3]. Further, the GRU network was upgraded by inspiration from the AM to consider more important contextual information.

Furthermore, factorization approaches were employed, and the preference score was computed by the dot product. Following the prediction scores, it is possible to recommend top-k POIs to a user, and there is a high chance that the user will go there if the score is higher. The Bayesian personalized ranking (BPR) framework [20] learned the parameters of GTCI-EAGRU. In the last stage, three general datasets were utilized to conduct extensive experiments. Five up-to-date POI recommendation methods were compared with Brightkite, Gowalla, and Foursquare to evaluate the model.

1-2- Main Contributions

1- The proposed architecture is presented by combining the development of the GRU model with the MF method, which aims to apply the strengths of the models and reduce the challenges of each of these methods. According to the MF method, in the CF approach, places visited on social networks by a user on social networks can affect the next POI of other users on those networks. However, CF-based approaches are weak in modeling sequential data and do not consider the effect of sequential data on modeling, while the next POI prediction problem is inherently a time sequence problem. Although recurrent models have been proposed for sequential data modeling, they have limitations. The traditional RNN model cannot integrate the corresponding check-in contextual information into the modeling. Newer recurrent models also consider the effect of temporal and spatial contextual information similarly, while they have a separate effect.

Therefore, there is a need to develop these models. In the recurrent layer of the proposed architecture, a development of the GRU model is presented using three attention gates that consider the contextual information separately and in terms of their importance.

2- Within the recurrent layer of the proposed architecture, the flexibility of the GRU model is employed, and the GRU model was expanded following the attention-based approach. Moreover, three additional attention gates were proposed, including timestamp contextual attention gate (Gts), geographical contextual attention gate (Gge), and temporal contextual attention gate (Gte). The Gts controls the influence of timestamp earlier visited locations, whereas Gge and Gte control the effect of the hidden state of the earlier recurrent unit based on geographical distances and time intervals between two successive check-ins, respectively. This innovation makes it possible to extend the model to another context.

3- In this research, user contextual information is classified into two categories of absolute and transitional content information. The first category includes check-in timestamp and geographical coordinates and the second one consists of the time interval and geographical distance between two successive consecutive check-ins. Our proposed architecture considers two types of absolute and transitional contextual information separately. This category focuses on developing a model to consider more contextual information in the future.

4. Some comprehensive experiments were conducted on three large-scale real-world datasets, namely, Brightkite, Gowalla [21], and Foursquare [14] that are widely used in related studies to predict the user POI in LBSNs. The aim was to show the effectiveness of the proposed GTCI-EAGRU architecture for the next POI recommendation.

1-3- Problem Statement

Human mobility prediction is important for a wide spectrum of LBSN applications, and the next POI recommendation is one of the usages of predicting people's mobility [1]. In some LSBNs, users share their location by registering check-ins. The check-ins gathered in LBSNs contain geographical and temporal contextual information (TCI), and each piece of information has a separate effect on predicting the user's next location [3, 8]. In previous studies, some restrictions were applied for dividing sequence into different check-in trajectories such as using the time interval of less than six hours [1]. Nonetheless, applying restrictions for the time interval and geographical distance, when considering registered check-ins in data preprocessing, is not a proper approach for the mentioned purpose. The AM can address the mentioned issue. Instead of using multiple assumptions to consider the time interval or geographical distance constraints between two check-ins, it can be addressed by automatic weighting given to the model inputs inspired by the AM.
According to evidence [2], CF-based approaches have weaknesses in sequential data modeling and fail to consider the effect of sequential data, while the problem of the next location prediction is inherently a matter of time sequence (Challenge 1). Traditional recurrent models are unable to consider contextual information, but this information is highly important in determining the next POI (Challenge 2). Meanwhile, some earlier studies, based on recurrent models, consider the effect of temporal and geographical contextual information (GCI) to be the same, while they have a different effect (Challenge 3). Furthermore, according to [11], some proposed architectures, which are a combination of recurrent models and AM, are highly complex (Challenge 4). In this work, the GTCI-EAGRU model was proposed to address the above-mentioned challenges.

1-4- Organizations

The remaining parts of this research are as follows: The related methods are briefly reviewed in Section 2. Sections 3 and 4 describe some preliminaries to the study and the details of the GTCI-EAGRU network, respectively. In Section 5, an illustration of the experiments is presented, followed by providing the results of the proposed method. Finally, Section 6 summarizes conclusions and an outline for future works.

2- Related Works

This section classifies related studies under three approaches generally used for the next POI recommendations, including CF, RNN, and AM. Table 1 provides a summary of related works with their challenges considered in our research.

Table 1. summarize of related works

Model Name	Model Approach	Method summery	challenges
[28] Unified method	CF based	Believing that time plays an important role in POI recommendations and defining a new problem, namely, the time-aware POI recommendation to recommend POIs for a given user at a specified time in a day	Focusing on temporal contextual information and paying less attention to geographic contextual information
[30] LORI	CF based	Applying a confidence coefficient for each user in the integration process and designing a learning-to-rank based algorithm to train confidence coefficients	Not taking into consideration time interval and geographical distance
[33] ST-RNN	RNN based	Extending RNN and using a transition matrix for capturing the temporal cyclic effect and geographical influence	Vanishing gradient problem in long sequence due to the use of the traditional RNN
[22] STGN	RNN based	Modifying the basic LSTM model slightly by introducing gates and cells to capture short- and long-term preferences	Considering the same effect for temporal and geographical contextual information
[8] SERM	RNN based	Jointly learning the embedding of multiple factors (user, location, time, and keywords) and the transition parameters of an RNN in a unified framework	Not taking into account the geographical distance in the training of this model
[35] CA-RNN	RNN based	Employing adaptive context-specific input matrices and adaptive context-specific transition matrices	Using a traditional RNN model and restrictions on paying attention to the contextual information, low performance
[1] ATST-LSTM	AM and RNN based	Developing an attention-based spatiotemporal LSTM network to focus on the relevant historical check-in records in a check-in sequence selectively using the spatiotemporal contextual information	Encountering with high complexity of implementation and a lack of attention to the scarcity
[6] Deep Move	AM and RNN based	Capturing complex dependencies and multi-level periodicity nature of humans using embedding, GRU, and AM	Not taking into account the time interval between two checks to model the behavioral pattern of user check-ins
[11] DAN-SNR	AM based	Makes use of the self-AM. By leveraging multi-head self-attention, the DAN-SNR can model long-range dependencies between any two historical check-ins efficiently and weigh their contributions to the next destination adaptively	Using only the attention mechanism and had low performance rather than applying recurrent neural networks for modeling the sequential influence and social influence

Note. CF: Collaborating filtering; RNN: Recurrent neural network; AM: Attention mechanism; LORI: Learning-to-rank-based integration; ST-RNN: Spatiotemporal-Recurrent neural network; STGN: Spatio-temporal gated network; SERM: Semantics-enriched recurrent model; CA-RNN: Context-Aware Recurrent Neural Networks; ATST-LSTM: Attention-based Spatiotemporal-Long short term memory; DAN-SNR: Deep attentive network for social-aware recommendation.

3- Preliminaries

The research problem is formulated, and the applied preliminaries in this study are presented in the following section.

3-1- Notations and Definitions

Table 2 presents some primary notations used in this study.

Definition 1 (Check-in): A check-in is an action that a user takes under a geographical and temporal context. In addition, it is a registration of a location in the LBSN that contains geographical and temporal information. When a user u checks in a location l (including latitude and longitude) with venue-Id v at the timestamp t, the check-in record can be modeled as a quadruple: cu, v, t < u, l, v, t >.

Definition 2 (Check-in sequence): A user's check-in sequence or S^u is a set of all user check-ins.

Definition 3 (Trajectory): Given a user u, a trajectory t is a sequence of chronologically ordered check-in associated with u. For example tr_u: $< u, l_1, v_1, t_1 >,\ldots,< u, l_i, v_i, t_i >,\ldots,< u, l_k, v_k, t_k >$, where tr_u is the trajectory of a user u before time tk. Here, a trajectory set $Tr^{(u)}$ is used to denote all the trajectories of user u.

Definition 4 (POI): In LBSNs, a POI is a spatial item related to a geographical location and known as a venue, including a hotel or an office. In this research, POI is represented by *v*, and the set of POIs is demonstrated as $V=\{v_1, v_2,...\}$. Each POI v has a unique identifier and geographical coordinate, consisting of geographical latitude and geographical longitude.

Definition 5 (the next POI recommendation): Given all users' trajectories, the aim of the next POI recommendation is to predict the most likely location vk that a user u will visit at a certain time point t_{N+1}.

Definition 6 (POI recommendations): Given a set of users' check-in sequences S^u and a set of POIs *V*, the POI recommendation task is to recommend top-k POIs that are preferable for user u.

Table 2: Notations and descriptions used in this study

Notations	*Descriptions*
u, l, v, & t	User, location (including latitude and longitude), venue or POI, and timestamp
$c_{u,v,\&\,t}$	A check-in recorded by user *u* in POI *v* and timestamp *t*
lat v & lng v	Latitude and longitude of POI *v* (i.e., geographical coordinates of POI *v*)
Δt & Δg	Time interval and geographical distance between two successive check-ins
S^u	A set of all check-ins generated by user *u*
Us, V, & T	Sets of users, POIs, and timestamp
v_τ^u	POI visited by user *u* at time step τ
g_τ^u & t_τ^u	Vector representations of geographical and temporal intervals
tr_u	A sequence of chronologically ordered check-ins related to *u*
$Tr^{(u)}$	All trajectories from user *u*
ϕu	The latent factor of user *u*
ϕv	The latent factor of POI *v*
ϕt	The latent factor of timestamp t
d	The number of latent dimensions
v^+	A set of positive POIs (visited venues) for each user $u \in Us$
v^-	A set of positive negative POIs (unvisited venues) for each user $u \in Us$
σ	Sigmoid function

3-2- MF in CF Based Approach

CF-based methods aim to discover similarities in the user's previous behavior and make predictions to the user based on a similar preference with other users [25]. There are various model-based CF algorithms, but MF is the most commonly applied in recommender systems [2]. MF seems to be the most accurate approach for lowering the problem from high levels of scarcity in the recommender systems database. Generally, MF models map both users and items to a joint latent factor space of dimensionality d in such a way that user-item interactions are modeled as inner products in that space. In the next POI recommendation, the item is the same POI or venue that a user has selected at the time of the check-in.

Accordingly, each venue v is related to a vector $q_v \in R^d$, and each user u is associated with a vector $p_u \in R^d$. For a given venue v, the elements of qv measure the extent to which the venue possesses those factors, positive or negative. For a given user u, the elements of p_u measure the extent of interest the user has in venues that are high on the corresponding factors, positive or negative. The resulting dot product ($q_v^T p_u$) captures the interaction between user u and venue v- the user u's overall interest in the venue's characteristics. This approximates the user's rating of venue v, which is denoted by r_{uv}, leading to the following estimate [36]:

$$\hat{r}_{uv} = p_u q_v^T \quad (1)$$

The objective is to minimize the prediction error or the loss function in Eq. (2) where K is the set of (u,v) pairs of known ratings [25, 36].

$$min \sum_{(u,v)\in k} (r_{uv} - p_u\ q_v^T) \quad (2)$$

Different approaches exist [13,36] for the extension of MF using RNN models to capture the user's dynamic preferences from the sequence of user's check-ins. Specifically, with respect to the sequence of a user's check-ins, the output of an RNN model can be effective in representing a user's dynamic preferences and modifying MF-based approaches.

3-3- GRU in RNN Based Approach

The next POI recommendation is immediately faced with the challenge of learning personalized user preferences for POIs and the sequential correlations jointly and efficiently between the check-ins [1]. To solve this problem, the RNN takes a sequence of inputs and learns the sequential pattern of the input sequence using hidden states [2,3,8]. The problem that the RNN faces is the exploding and vanishing gradients; therefore, it cannot capture long-term preferences [2,11]. The problems can be solved by long-short term memory (LSTM), which employs a gate mechanism and can capture long-term preferences [1,23].
The information flow among consecutive LSTM cells is controlled through input, forget, and output gates. LSTM resolves the problems of the RNN, but it has three gates thus the training of an LSTM-based model is slower and requires a large amount of training data. GRU [4,6,23,40] has updated and reset gates in the network, dealing with the update degree of each hidden state. In fact, it determines which information should pass to the next state [2,3]. Fig. 1 displays the block diagram of basic GRU. As shown, GRU uses only two gates (i.e., reset and update gates). The GRU-based model can be trained faster and perform better compared to LSTM when there are less training data. GRU calculates hidden state h_τ at time τ from the output of update gate z_τ, reset gate r_τ, current input x_τ, and previous hidden state $h_{\tau-1}$. $\hat{h}_\tau$ and h_τ are computed from the reset gate as follows:

$$z_\tau = \sigma (W_z x_\tau + U_z h_{\tau-1} + b_z) \quad (3)$$

$$r_\tau = \sigma (W_r x_\tau + U_r h_{\tau-1} + b_r) \quad (4)$$

$$\hat{h}_\tau = tanh (W x_\tau + U (r_\tau \odot h_{\tau-1}) + b_h) \quad (5)$$

$$h_\tau = (1 - z_\tau) \odot h_{\tau-1} + z_\tau \odot \hat{h}_\tau \quad (6)$$

where $\odot$ is a basic multiplication operation, and W and U represent weight matrices for training the network.

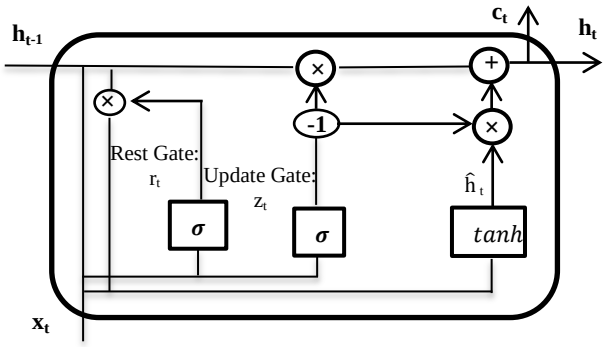


Fig. 1. An illustration of the GRU cell including two gates.[3]

3-4- Attention Mechanism

The AM was proposed based on the selective AM in the human visual system [1,18]. It should be noted that humans are prone to giving higher attention to key parts of the input, helping in breaking down a complex input into simpler parts that can easily be processed accordingly. Rather than paying attention to all available information, selective attention mainly focuses on the most relevant information in a system. Accordingly, learning to pay attention to the specific components of the input data resulted in different attention models in deep learning [4,11].
The present study proposed a novel model that is applied to this mechanism for the next location prediction. The key idea in the AM is that inputs are mapped to query, key, and value vectors. The outputs are calculated by taking the weighted sum of the value vectors where weights are determined by a function of query and key values [11]. Specifically, the attention function presents a query and a group of key-value pairs to a context vector, which is a weighted sum of all values. The queries, keys, and values are merged as Q, K, and $Vval$ matrices, respectively [1,18]. For the output of the attention function, an alignment function or the compatibility function, which measures the quality of the match between the input query matches and the corresponding key, calculates the weight assigned to each value. Eq. (7) is used for the computation of the matrix of outputs where (Q, K) refers to the attention function [18]:

$$Attention\ (Q, K, V_{val}) = Softmax\ (f(Q,K))V_{val} \quad (7)$$

Additive attention and dot-product (multiplicative) attention are two of the most commonly used attention functions and are defined as follows [1]:

$$f_{add}(Q,k) = tanh(w_Q + W_K\ k) \quad (8)$$

$$f_{mul}(Q,k) = Q\ k^T \quad (9)$$

In theory, these two functions are similar in computation complexity. Additive attention and dot-product attention use a feed-forward neural network with a single hidden layer for the calculation and optimized matrix multiplication operation, respectively [1]. The present project, which is inspired by previous studies [2,3], employed a feed-forward neural network to calculate the alignment function to develop the GRU model.

4- Proposed GTCI-EAGRU Model Description

The GTCI-EAGRU architecture consists of input, embedding, recurrent, and output layers. Fig. 2 presents a schematic of our purpose architecture. The details of these layers and the learning procedure for the parameters are provided as follows:

4-1- Input Layer

The input layer contains model inputs that include absolute context and relative or transition context. In the proposed model, the absolute context is user id, timestamp, geographical coordinates (including latitude and longitude), and venue id. Further, the relative context (also called the transition context) is the time interval (Δt) and geographical distance (Δg) between two successive check-ins.

The geographical distance and time intervals are calculated in the input layer. For a given user u, venue v_n, and time t^τ , the geographical distance (Δg^τ) and time interval (Δt^τ) between the POIs at current time t^τ and previous time $t^{\tau-1}$, as well as the given venue v_n and venue v_{n-1} previously visited at time are computed as: $\Delta t^\tau = t^\tau - t^{\tau-1}$ and Δg^τ = dist (lat v_1, lng v_1, lat v_2, lng v_2), respectively, where dist () is the Haversine and its function is as Eq.(10)[1]. It should be noted that the Haversine distance is the angular distance between two points on the surface of a sphere. The former coordinate of each point is taken as the latitude and the latter one is the longitude given in radians. The data dimension must be two[2].

$$D(x,y)=2arcsin\left[\sqrt{sin^2\left(\frac{x_1-y_1}{2}\right)+cos(x_1)cos(y_1)sin^2\left(\frac{x_2-y_2}{2}\right)}\right] \quad (10)$$

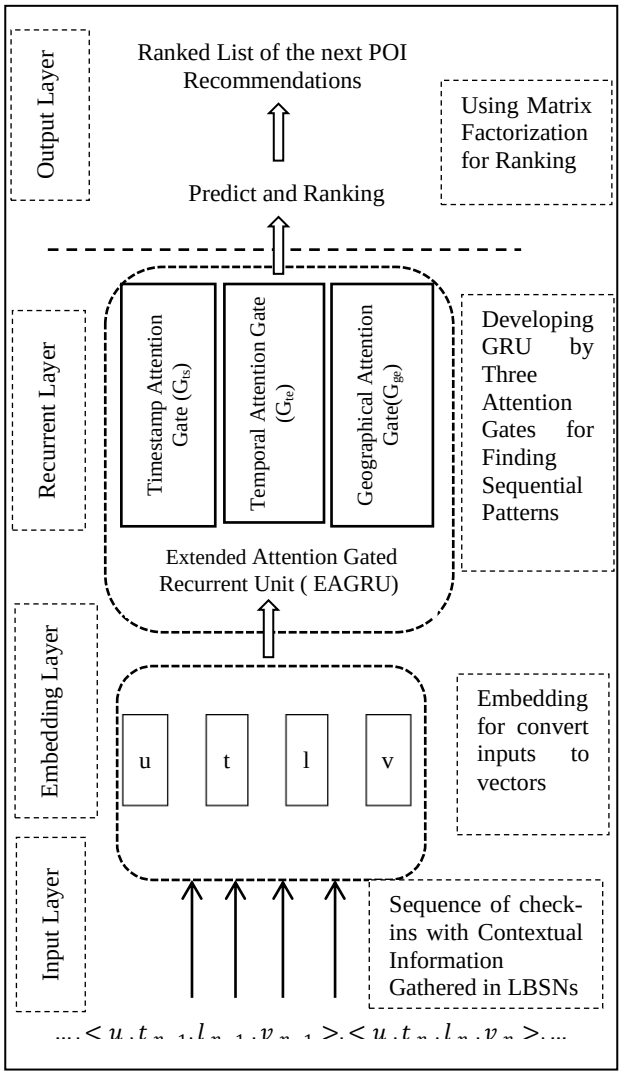


Fig. 2. GTCI- EAGRU architecture with input, embedding, recurrent, and output layers

Note. GTCI- EAGRU: Geographical temporal contextual information Extended attention gated recurrent unit; LBSN: Location-based social network. In the recurrent layer, the GRU model is extended with three additional attention gates, and a ranked list of the next POI recommendations is provided in the output layer.

4-2- Embedding Layer

This layer is for embedding inputs from the check-in sequence before it goes to the recurrent layer. In this layer, embedding or latent factors are generated from the inputs. In addition, the latent factors of the user, namely, POI (or venue) and time are generated as $\emptyset u_i \in U$, $\emptyset v^\tau_i \in V$ and time $\emptyset t^\tau \in$ T, respectively. Note that θe = {U, V, T} denotes the set of the parameters of the embedding layer. Next, the latent factors of venue $\emptyset v^\tau_j$, the latent factors of the given time $\emptyset^\tau_t$, and the contextual transition features (Δg_τ and Δt_τ) are passed to the recurrent layer for training using GTCI-EAGRU.

4-3- Recurrent Layer

In this layer, the GRU model was developed with three attention gates. Following Manotumruksa et al. [2] and Kala et al. [3], this study presented timestamp attention gate (G_{ts}), geographical attention gate (G_{ge}), and temporal attention gate (G_{te}). The input of G_{ts} is the check-in time (i.e., the time that the check-in is registered by a user in LBSN and includes the year, month, day, hour, minute, and second). This gate is used to specify a more important timestamp in the sequence of historical check-ins of a user. However, the input of G_{te} is the time interval (Δt) between two successive check-ins used to specify more important time intervals in the sequence of historical check-ins of a user. The input of G_{ge} is the geographical distance (Δg) between two successive check-ins applied to specify a more important geographical distance in the sequence of historical check-ins of a user. The output of this layer is the hidden state of the recurrent unit at time step τ, h_τ, and is defined as Eq. (11):

$$h_\tau = f(\emptyset v_j^\tau, \emptyset t^\tau, \Delta t^\tau, \Delta g^\tau; \theta_r) \quad (11)$$

As mentioned earlier, the purposed model treats the absolute and relevant (or) transition contextual information separately. It is noteworthy that this contextual information has a different effect on the user's dynamic preference and requires independent consideration. The following part describes the extension of the traditional GRU for the integration of absolute and relevant contextual information.

Generally, in the GRU model, given the user's sequence of check-ins S^u and dynamic preference at time step τ, the hidden state (h^τ) is estimated by the update and reset gates, which are defined as:

[1] https://scikit-learn.org.

[2] In general, the Eq. (10) is used to calculate the Haversine distance between samples in X and Y (x_1 and x_2 are latitude and longitude of X and y_1 and y_2 are latitude and longitude of Y, respectively).

$$z_\tau = \sigma(W_z \emptyset v_j^\tau + U_z h_{\tau-1} + b_z) \quad (12)$$

$$r_\tau = \sigma(W_r \emptyset v_j^\tau + U_r h_{\tau-1} + b_r) \quad (13)$$

$$\hat{h}_\tau = tanh(W_h \emptyset v_j^\tau + U_h (r_\tau \odot h_{\tau-1}) + b_h) \quad (14)$$

$$h_\tau = (1 - z_\tau) \odot h_{\tau-1} + z_\tau \odot \hat{h}_\tau \quad (15)$$

where $\emptyset v_j^\tau$ represents the latent factor of venue j that user i visited at time step τ.

$\sigma()$ and *tanh()* are the sigmoid and hyperbolic tangent functions, respectively. Furthermore, U is a recurrent connection weight matrix that captures sequential signals between every two adjacent hidden states h_τ and $h_{\tau-1}$ using $\odot$, which shows the element-wise product. Moreover, W and b are the transition matrix between the latent factors of venues and the corresponding bias, respectively. It should be noted that $\theta_r = \{W, U, b\}$ denotes the set of the parameters of the recurrent layer. Overall, W is the transition matrix between the latent factors of venues and b indicates the corresponding bias. Additionally, U is a recurrent connection weight matrix that captures sequential signals between every two adjacent hidden states. All the recurrent layer parameters (i.e., W_z, U_z, and b_z) are the set of the parameters of the update gate. W_r, U_r, and b_r, as well as W_h, $U_{h,}$ and b_h are the set of parameters of the reset gate and candidate hidden state, respectively. Similarly, W_{Gts}, U_{Gts}, and b_{Gts} are the set of the parameters of our proposed G_{ts}. Finally, W_{Gge}, $U_{Gge,}$ and b_{Gge}, as well as W_{Gte}, $U_{Gte,}$ and b_{Gte} are the set of the parameters of our proposed G_{ge} and G_{te}, respectively.

At current step τ, the correlation between the latent factor of absolute contexts ϕt^τ and the hidden state from the earlier step $h_{\tau-1}$ is calculated by Eq. (16):

$$G_{ts} = \sigma(W_{Gts} h_{\tau-1} + W_{Gts} \phi t^\tau + b_{Gts}) \quad (16)$$

To effectively model the users' sequential order of check-ins, the relevant contextual information needs to be examined separately. To address this issue, the current study proposed G_{ge} and G_{te} to individually incorporate the geographical distance (Δg_τ) and time interval (Δt_τ) between two check-ins as Eqs. (17) and (18):

$$G_{ge} = \sigma(W_{Gge} h_{\tau-1} + W_{Gge} \Delta g^\tau + b_{ge} \quad (17)$$

$$G_{te} = \sigma(W_{Gte} h_{\tau-1} + W_{Gte} \Delta t^\tau + b_{te}) \quad (18)$$

With the proposed gates for GTCI-EAGRU architecture, the equations of the traditional GRU are updated as Eqs. (19), (20), and (21):

$$z_\tau = \sigma(W_z \emptyset v_j^\tau + U_z h_{\tau-1} + \boldsymbol{W_z((G_{ts} \odot \phi t^\tau) + (G_{ge} \odot \Delta g^\tau) + (G_{te} \odot \Delta t^\tau))} + b_z) \quad (19)$$

$$r_\tau = \sigma(W_r \emptyset v_j^\tau + U_r h_{\tau-1} + \boldsymbol{W_r((G_{ts} \odot \phi t^\tau) + (G_{ge} \odot \Delta g^\tau) + (G_{te} \odot \Delta t^\tau))} + b_r) \quad (20)$$

$$\hat{h}_\tau = tanh(W \emptyset v_j^\tau + U (r_\tau \odot h_{\tau-1}) + \boldsymbol{W_r((G_{ts} \odot \phi t^\tau) + (G_{ge} \odot \Delta g^\tau) + (G_{te} \odot \Delta t^\tau))} + b_h) \quad (21)$$

In the following section, the hidden sate h_τ will be updated and as previously mentioned, it will be the output of the recurrent unit at time step τ.

4-4- Output Layer

In the next POI recommendations based on the MF approach, recommendations are mainly derived from a dot product of the latent factors of users $U \in R^{|U| \times d}$ and venues $V \in R^{|V| \times d}$ where d is the number of latent dimensions (i.e. $\hat{c}_{i,j} = \emptyset u_i \emptyset v_j^T$) and $\emptyset u_i$ and $\emptyset v_j$ denote the latent factors of user i and venue j, respectively [2,36]. In the output layer, the preference of user u on venue v at timestamp t is estimated using Eq. (22):

$$\hat{c}_{u,v,t} = \emptyset u_u h^{\tau T} \quad (22)$$

According to previous works, the pairwise loss function outperformed the classification loss function in learning patterns from sequential data and was more efficient for the network training of the recurrent-based recommendation [2,3,13,20]. Therefore, following Manotumruksa et al. [2,13], the pairwise BPR [20] can be applied to estimate the embedding and recurrent layer parameters and the probability distribution over all venues given the hidden state h^τ.

4-5- Network training

This study employed datasets consisting of a set of sampled triplets each containing one user and a pair of POIs in which one POI is positive (known as visited) while the other one is negative (known as unvisited). As mentioned earlier, this study applied the pairwise BPR to learn the embedding and recurrent layer parameters ($\Theta = \{\theta_e, \theta_r\}$). Based on an underlying assumption, stating that a user prefers the observed POI to all unobserved ones, BPR considers the relative order of the predictions for the pairs of POIs [1,4]. At each sequential position k in the BPR framework, the goal of GTCI-EAGRU is to maximize the following probability [1,4,20]:

$$P(u, t, v > v') = g(o_{u,t,v} - o_{u,t,v'}) \quad (23)$$

v and v' stand for a positive (visited) POI and a negative (unvisited) POI, respectively, and $g(\cdot)$ represents a nonlinear function defined by Eq. (24) as [1, 20]:

$$g(x) = \frac{1}{1+e^{-x}} \quad (24)$$

The objective function of the network for the next POI recommendation can be solved by integrating the loss function and a regularization term as follows [20]:

$$J = -\sum_{(v,v')} \ln P(u, t, v > v') + \lambda/2 \, \| \Theta \|^2 \quad (25)$$

where λ is used to specify the power of regularization and Θ is the parameter set. The dimension of the latent factors d and hidden layers h_τ of GTCI-EAGRU architecture d = 10 across three datasets can be set based on methods by Manotumruksa et al. [2] and Kala et al. [3], and all

embedding and recurrent layers' parameters can be randomly be initiated with a Gaussian distribution. Initially, the learning rate and the batch size are set to 0.001 and 256, respectively. An Adam optimizer was employed to optimize the model parameters. The output of the GTCI-EAGRU model is a set of scores for POIs, similar to their likelihood of being the next POI in each sequence. A summary of the learning algorithm of GTCI-EAGRU is provided as follows:

Algorithm 1: Training of GTCI-EAGRU	
Input: Set of users Us and set of historical check-in sequences S^u	
Output: GTCI-EAGRU model $\{\Theta\}$	
//construct training instances	
1.	Initialize $D=Us_u$ $D^u = \emptyset$ D^u is a set of check-in trajectory samples combined with negative POIs of u
2.	**For** each user $u \in U$s **do**
3.	**For** each check-in sequence $S^u = \{s_{t1}{}^u, s_{t2}{}^u, \ldots, s_{tn}{}^u\}$ **do**
4.	Get the set of negative samples v'
5.	**For** each check-in activity in S^u **do**
6.	Compute the embedded vector $v_\tau{}^u$
7.	Compute the geographical contexts vector $g_\tau{}^u$
8.	Compute the temporal contexts vector $t_\tau{}^u$
9.	**End for**
10.	Add a training instance $(\{v_\tau{}^u, g_\tau{}^u, t_\tau{}^u\}, \{v'\})$ into D^u
11.	**End for**
12.	**End for**
//train the model	
13.	Initialize the parameter set Θ
14.	**While** (exceed(maximum number of iterations)==FALSE) **do**
15.	**For** each user u in U **do**
16.	Randomly select a batch of instances $D_b{}^u$ from D^u
17.	Find Θ minimizing the objective (23) with $D_b{}^u$
18.	**End for**
19.	**End While**
20.	Return the set of parameter Θ

5- Experimental Result and Analysis

This section presents the experimental setup and empirical results of this study. Empirical experiments are conducted on three public datasets in LBSNs for validating the efficiency of the proposed method. To address the challenges made in Section 2-4, the experiments are designed for the following research questions:

RQ1: How can the basic GRU architecture be extended to separately consider the absolute and relative (or transition) contextual information associated with the sequence of check-ins?

RQ2: Is it important to model absolute and relative (or transition) contextual information separately?

RQ3: Does GTCI-EAGRU that leverages multiple types of contextual information improve prediction accuracy by applying additional attention gates? Or, does it outperform the previous methods?

5-1- Datasets and Experimental Settings

The experiments were conducted for evaluating three publicly LBSN datasets (i.e., BrightKite[1], Gowalla[2], and Foursquare[3] datasets). Following Manotumruksa et al. [2] some deletions were made to lessen data sparsity and cold start problems. Users with less than 10 check-ins and POIs with less than 10 were eliminated from the three datasets. Table 3 presents an overview of the statistics of the three datasets. In this study, a check-in record is a quadruple composed of a user, the corresponding check-in timestamp, the geographical coordinates of the check-in, and a location Id or POI. The check-in records in these three datasets were regarded as user sequences. The density calculation formula for three datasets is as follows [38]:

$$Density = \frac{|check-ins|}{|users| \times |POIs|} \quad (26)$$

Table 3. Statistics of the three datasets

Dataset	*#Users*	*#Check-ins*	*#POIs*	*Density*
Brightkite	915	676721	7527	0.0982
Gowalla	1047	614340	5011	0.1170
Foursqure	615	108195	19245	0.0091

A leave-one-out evaluation method was adopted to evaluate the efficiency of the proposed GTCI-EAGRU architecture based on earlier works [2], [3]. Each user's most recent check-in was taken as the base, and 100 POIs, which had not been visited before, were randomly selected for this purpose. They were the testing set, and the other remaining check-ins were considered as the training set. The task of the GTCI-EAGRU was to rank those 100 venues for each user as their preferred contexts (i.e., timestamp, time interval, and geographical distance), aiming at ranking highest the recent, ground truth check-in. Following Manotumruksa et al. [2] and Kala et al. [3], the researchers set the dimension of the latent factors d and hidden layers hτ of the proposed GTCI-EAGRU architecture: d = 10. As mentioned before, Gaussian distribution [32] was employed for the random initialization of the recurrent layer's parameters, and Adam Optimizer [39] was utilized for optimizing the parameters because it had a faster convergence compared to the stochastic gradient descent optimization, which automatically adjusts the learning rate for each iteration. In addition, the batch size and the dropout rate were set to 256 and 0.2, respectively, to prevent overfitting.

5-2- Comparison

The following five up-to-date methods were compared to validate the efficiency of the GTCI-EAGRU in the next POI recommendation task. Table 4 summarizes these methods into different aspects. Based on data, they are categorized into MF-

[1] https://snap.stanford.edu/data/loc-brightkite.html
[2] https://snap.stanford.edu/data/loc-gowalla.html
[3] https://sites.google.com/site/yangdingqi/home/foursquare-dataset

, RNN-, and AM-based approaches. The compared models are also classified according to the use of GCI and TCI.
A brief description of these models is given below:
STGN: Spatio-temporal gated network was proposed by Zhao et al. [22], and improved the LSTM network, in which STGs are introduced for capturing the Spatio-temporal relationships between successive check-ins. By introducing new gates and cells to capture short- and long-term preferences, STGN modified the basic LSTM model.
ARNN: An attentional RNN was proposed by Guo et al. [7] to jointly model the transition regularities and sequential regularity of similar locations (neighbors). Using embedding, knowledge graph, LSTM, and AM, the ARNN captured sequential, spatial, temporal, and semantic influences.
GeoSAN: By introducing a new loss function, Lian et al. [37] resolved the sparsity issue. **GeoSAN** represents the hierarchical gridding of each GPS point with a self-attention based geography encoder for better use of geographical information.
DRCF: To benefit from the traditional RNN to model the sequential order of users' check-ins, Manotumruksa et al. [13] extended NeuMF. DRCF has two components each having its recurrent layer.
CARA: By employing embedding, GRU, and two gating mechanisms, Manotumruksa et al. [2] captured various types of the impact of different contextual information.
Following earlier works [6-8,22], the current study used prediction accuracy (*Acc@k*, $k = 10$) for evaluating the performance of the above-mentioned methods and checking if the ground-truth location can be found in the top-*k* recommendation list. Generally, the *Accuracy@* is defined by Eq. (27) as follows [29]:

$$Accuracy@k = \frac{number\ of\ samples\ correctly\ predicted}{total\ number\ of\ samples} \quad (27)$$

Table 4. Summary of all the baseline methods used in this study

Methods	*Approaches and Contextual Information*				
	MF	***RNN***	***AM***	***GCI***	***TCI***
STGN	×	√	×	√	√
ARNN	×	√	√	√	√
GeoSAN	×	×	√	√	√
DRCF	√	√	×	×	×
CARA	√	√	√	√	√
GTCI-EAGRU	√	√	√	√	√

Note. MF; Matrix factorization; RNN: Recurrent neural network; AM: Attention mechanism; GCI: Geographical contextual information; TCI: Temporal contextual information; STGN: Spatio-temporal gated network; ARNN: Attentional Recurrent Neural Network; GeoSAN: Geography-aware sequential recommender based on the Self-Attention Network; DRCF: Deep Recurrent Collaborative Filtering; CARA: Contextual attention recurrent architecture ; GTCI-EAGRU: Geographical temporal contextual information-extended attention gated recurrent unit.

5-3- Results and Discussion

Table 5 compares the recommendation results of six methods on the three datasets. The numbers in bold in each column represent the best performance.

Table 5. Comparison of different methods in recommendation performance

Methods	Acc@10		
	Brightkite	**Gowalla**	**Foursquare**
STGN	0.2020	0.5231	0.3017
ARNN	-	0.2336	0.4285
GeoSAN	0.6425	0.6028	0.4867
DRCF	0.7363	-	0.8805
CARA	0.7385	-	0.8851
GTCI-EAGRU	**0.9751**	**0.9606**	**0.8901**

Note.: STGN: Spatio-temporal gated network; ARNN: Attentional Recurrent Neural Network; GeoSAN: Geography-aware sequential recommender based on the Self-Attention Network; DRCF: Deep Recurrent Collaborative Filtering; CARA: Contextual attention recurrent architecture; GTCI-EAGRU: Geographical temporal contextual information-extended attention gated recurrent unit.

The comparison of the experimental results of the models demonstrated that the use of AM alone (i.e., the GeoSAN model) has not increased prediction accuracy. Moreover, the experimental results of other previous studies (e.g., DAN-SNR) revealed lower evaluation metrics values. Although the STGN model separately considered the GCI and TCI, it did not use the attenuation mechanism approach. It applied the LSTM model and was less prediction accurate compared to models that employed the GRU such as CARA. Although the ARNN model applied the LSTM model, it had a higher accuracy prediction in comparison with the STGN model due to the use of the attenuation mechanism. Similar to the STGN, it had less prediction accuracy compared to models that considered the GRU model.
The GeoSAN model only uses the AM for location recommendation, and despite considering geographical and TCI, it is less prediction accurate than DRCF and CARA models. The DRCF model pays attention to the sequence of previously visited venues while not taking into consideration the contextual information related to the check-ins. Thus, its prediction accuracy is lower than that of the hybrid models. However, it should be stated that the performance of these hybrid approaches was not worse than that of RNN and LSTM. Thus, it is worth modeling geographical and spatial contextual information for the task of the next POI recommendations. It means that it is insufficient to have a good network architecture, but more geographical and spatial contextual information of human check-in behaviors should be taken into account to obtain excellent results [1]. This is the reason for the outperformance of CARA over DRCF.
The accuracy prediction in the CARA model is higher compared to other models due to the separate use of TCI and GCI and a combination of the RNN, attention, and factoring approaches. Inspired by the idea behind this model, the researchers introduced a new initiative to employ three gates in the GRU model to address GCI and TCI to better predict the accuracy of the next POI recommendation. As mentioned in previous sections, the proposed model uses three separate attention gates, namely, G_{ts}, G_{ge}, and G_{te}, which consider the timestamp, geographical distance, and time interval between successive check-ins, respectively, and the output of each of

them separately affects the values of the reset and update gates of the GRU model. As depicted in Fig. 3, the experiment results of the proposed models indicate that it has achieved this goal, and the accuracy prediction has been improved in the proposed GTCI-EAGRU architecture.

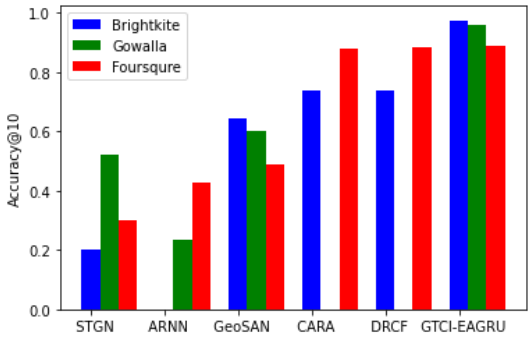


Fig. 3. Comparison of GTCI-EAGRU with baseline methods in terms of Accuracy@10 on three datasets

To answer RQ1 for the development of the GRU model, it should be mentioned that three gates were introduced and implemented as a feed-forward network. The output of these gates affects the values of the GRU reset and update gates, and they are responsible for controlling the geographical and temporal information of the user's trajectory data. To answer RQ2 and RQ3, these results were obtained (Table 6) by comparing the accuracy prediction of the GTCI-EAGRU model with up-to-date architectures.

Table 6. Percentage of Improvement of GTCI-EAGRU

Methods	*Percentage of Improvement*		
	Brightkite (%)	***Gowalla (%)***	***Foursquare (%)***
STGN	79.28	45.54	66.10
ARNN	-	75.68	51.86
GeoSAN	34.11	37.25	45.32
DRCF	24.49	-	01.08
CARA	24.26	-	00.56
Improvement	40.54	52.82	32.98
On Average in three Dataset		42.11	

6- Conclusions

In recent years, the next POI recommendation is of great importance for a wide spectrum of LBSN applications. The influences of contextual information (e.g., spatial and temporal context information) are crucial for analyzing individual behaviors for personalized POI recommendations. Hence, many studies have considered this contextual information to improve the performance of POI recommendation algorithms such as the CF and RNN. There are still many challenges regarding how to integrate contextual information to accurately model users' complex behavioral preferences and recommend reliable POIs to users.

The current study proposed a novel GTCI-EAGRU for the next POI recommendation by addressing the challenges concerning previous studies. Our proposed architecture was presented with the development of the GRU model, in which the contextual information of the user trajectory data is considered separately. Moreover, the development of the model inspired by the AM makes contextual information more important in modeling sequential user data. POIs were scored to provide recommendations to a user from her/his historical check-ins. The simple development of this model for considering more contextual information is one of the other features of the proposed model.

By comparing the experimental results of baseline methods, an increase in the accuracy of prediction indicates the importance of considering contextual information separately. The proposed GTCI-EAGRU architecture with three additional contextual attention gates worked well for the next POI recommendation.

In this study, the comprehensive experiments conducted on three large- scale datasets from the Brightkite, Gowalla, and Foursquare demonstrated a significant improvement in the GTCI-EAGRU architecture for the next POI recommendations compared with various up-to-date recurrent architectures and many different recent factorization approaches.

To enhance the quality of recommendations for the next POI, the GTCI-EAGRU architecture could be enriched by adding the impact of each user's social relationships with other users on LBSNs. Furthermore, it can be possible to include more contextual information (e.g., visual and text information) related to users' check-ins or the weather condition of the check-in registration location as well.

Acknowledgment

This manuscript is prepared based on PhD. thesis of the first author at Rasht Branch, Islamic Azad University, Rasht, Iran.

References

[1] L. Huang, Y. Ma, Sh. Wang, Y. Liu, "An Attention-based Spatiotemporal LSTM Network for Next POI Recommendation", Journal of IEEE Transactions on Services Computing, vol. 12, 2019, pp. 1-13.

[2] J. Manotumruksa, C. Macdonald, I. Ounis, "A Contextual Attention Recurrent Architecture for Context- Aware Venue Recommendation", in 18th ACM SIGIR Conference on Research and Development in Information Retrieval, 2018, pp. 555-564.

[3] K. Kala, M. Nandhini," Context Category Specific sequence aware Point of Interest Recommender System with Multi Gated Recurrent Unit", Journal of Ambient Intelligence and Humanized Computing, 2019, https://doi.org/10.1007/s12652-019-01583-w.

[4] C. Liu, J. Liu, J. Wang, S. Xu, H. Han, Y. Chen, "An Attention-Based Spatiotemporal Gated Recurrent Unit Network for Point-of-Interest Recommendation", International Journal of Geo-Information, vol. 8, No. 8, 2019, pp.355-373.

[5] S. Wang, Z. Bao, J. Culpepper, G. Cong, "A Survey on Trajectory Data Management, Analytics, and Learning", ACM Computing Surveys, vol. 54, No.3, 2020, pp. 1-33.

[6] J. Feng Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, D. Jin, "DeepMove: Predicting Human Mobility with Attentional Recurrent Networks", in 18th ACM IW3C2 Conference on International World Wide Web, 2018, pp. 1459-1468.

[7] Q. Guo, Z .Sun, J. Zhang, Y. Theng, "An Attentional Recurrent Neural Network for Personalized Next Location Recommendation", in 34th ACM AAAI Conference on Artificial Intelligence, 2020, pp. 83-90.

[8] D.Yao,C. Zhang, J. Huang, J. Bi, "SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories", in 17th ACM CIKM Conference on Information and Knowledge Management, 2017 ,pp. 2411-2414.

[9] L. Zhang, Z. Sun, J. Zhang, H. Kloeden, F. Klanner, "Modeling hierarchical category transition for next POI recommendation with uncertain check-ins", Journal of Information Sciences, Elsevier, vol.515, 2019, pp. 169-190.

[10] L. Chang, W. Chen, J. Huang, Ch. Bin, W. Wang, "Exploiting multi-attention network with contextual influence for point-of-interest recommendation", Journal of Applied Intelligence, vol. 51, 2021, pp. 1904–1917.

[11] L. Huang, Y. Ma, Y. Liu, K. He, "DAN-SNR: A Deep Attentive Network for Social-Aware Next Point-of-Interest Recommendation", Journal of ACM Transactions on Internet Technology, Vol.21, No.2, 2020, pp. 1–27.

[12] G. Christoforidis, P. Kefalas, A. Papadopoulos, Y. Manolopoulos, "RELINE: Point-of-Interest Recommendations using Multiple Network Embeddings", Journal of Knowledge and Information Systems, Vol. 63, No.4, 2019, pp. 791-817.

[13] J. Manotumruksa, C. Macdonald, I. Ounis, "A Deep Recurrent Collaborative Filtering Framework for Venue Recommendation", in 17th ACM CIKM Conference on Information and Knowledge Management, 2017, pp. 1429-1438.

[14] D. Yang, D Zhang, V. Zheng, Z. Yu, "Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs", Journal of IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol.45, No.1, 2014, pp. 129 – 142.

[15] M. Quadrana, P. Cremonesi, D. Jannach, "Sequence-Aware Recommender Systems", Journal of ACM Computing Surveys, Vol.51, No.4, 201, pp. 1–36.

[16] Q. Cui, Y. Tang, S. Wu, L. Wang, "Distance2Pre: Personalized Spatial Preference for Next Point-of-Interest Prediction", in PAKDD Conference on Knowledge Discovery and Data Mining, 2019, pp. 289-301.

[17] Q. Gao, F. Zhou, G. Trajcevski, K. Zhang, T. Zhong, F. Zhang, "Predicting Human Mobility via Variational Attention", in IW3C2 Conference on International World Wide Web Conference Committee, 2019, pp. 2750–2756.

[18] A.Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need", in 31th NIPS Conference on Neural Information Processing System, 2017, pp. 5998-6008.

[19] Y. Chen, C. Long, G.Cong, C. Li, "Context-aware Deep Model for Joint Mobility and Time Prediction",in 13th ACM WSDM Conference on Web Search and Data Mining, 2020, pp. 106-114.

[20] S. Rendle, C. Freudenthaler, Z. Gantner, L. Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback", in 25th ACM UAI Conference on Uncertainty in Artificial Intelligence, 2009, pp. 452–461.

[21] E.Cho, S. Myers, J. Leskovec, "Friendship and Mobility: User Movement in Location-Based Social Networks", in 17th ACM KDD Conference on Knowledge Discovery and Data Mining, 2011, pp. 1082–1090.

[22] P. Zhao, H. Zhu, Y. Liu, J. Xu, F. Zhuang, V. Sheng, X. Zhou, "Where to Go Next: A Spatio-Temporal Gated Network for Next POI Recommendation",. in 33th AAAI Conference on Artificial Intelligence, 2019, pp. 5877-5884.

[23] A M. Islam, M. M. Mohammad, S. S. Das, M. E. Ali, "A Survey on Deep Learning Based Point-Of-Interest (POI) Recommendations", 2020, arXiv:2011.10187v1.

[24] C. Zheng, D. Tao, "Attention-Based Dynamic Preference Model for Next Point-of-Interest Recommendation". in 15th Springer WASA Conference on Wireless Algorithms, Systems, and Applications, 2020, pp. 768–780.

[25] D. K. Bokde, Sh. Girase, D. Mukhopadhyay, "Role of Matrix Factorization Model in Collaborative Filtering Algorithm: A Survey", International Journal of Advance Foundation and Research in Computer, vol.1, 2014, pp. 111-118.

[26] M. Gan, L. Gao, "Discovering Memory-Based Preferences for POI Recommendation in Location-Based Social Networks", International Journal of Geo-Information (IJGI), Vol.8, No.6, 2019, pp. 279-294.

[27] X. Meng, J. Fang, "A Diverse and Personalized POI Recommendation Approach by Integrating Geo-Social Embedding Relations", Journal of IEEE Access, Vol.8, 2020, pp. 226309- 226323.

[28] Q. Yuan, G. Cong, Z. Ma, A. Sun, N. Thalmann, "Time-aware Point-of-interest Recommendation", in 36th ACM SIGIR Conference on Research and development in Information Retrieval, 2013, pp. 363–372.

[29] P. Wang, H. Wang, H. Zhang, F. Lu, S. Wu, "A Hybrid Markov and LSTM Model for Indoor Location Prediction", Journal of IEEE Access, Vol.7, 2019, pp. 185928 – 185940.

[30] J. Li, G. Liu, C. Yan, C. Jiang, "LORI: A Learning-to-Rank-Based Integration Method of Location Recommendation", IEEE Transactions on Computational Social Systems, Vol.6, No.3, 2019, pp. 430 – 440.

[31] L.Yao, Q. Z. Sheng, Y. Qin, X. Wang, A. Shemshadi, Q. He, "Context-aware Point-of-Interest Recommendation Using Tensor Factorization with Social Regularization", in 38th.ACM SIGIR Conference on Research and Development in Information Retrieval, 2015, pp. 1007–1010.

[32] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T. Chua, "Neural collaborative filtering", in 26th ACM IW3C2Conference on World Wide Web Conference Committee, 2017, pp. 173-182.

[33] Q. Liu, S.Wu, L. Wang, T. Tan, "Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts", in 30th ACM AAAI Conference, 2016 , pp. 194–200.

[34] S. Kumar, M.I. Nezhurina, "An ensemble classification approach for prediction of user's next location based on Twitter data", Journal of Ambient Intelligence and Humanized Computing, Vol.10, No. 11, 2018, pp. 4503-4513.

[35] Q. Liu, S. Wu, D. Wang, Z. Li, L. Wang, "Context-Aware Sequential Recommendation", in ICDM Conference on Data Mining, IEEE, 2016, pp. 1053-1058.

[36] D. Bokde, S. Girase, D. Mukhopadhya, "Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey", Procedia Computer Science, Vol.49, 2015, pp. 136-146.

[37] D. Lian, Y. Wu, Y. Ge, X. Xie, E. Chen, "Geography-Aware Sequential Location Recommendation", in 26th ACM

ICGKDD Conference on Knowledge Discovery and Data Mining, 2020, pp. 2009–2019.

[38] K. Yang, J. Zhu, "Next POI Recommendation via Graph Embedding Representation from H-Deepwalk on Hybrid Network", Journal of IEEE Access, Vol 7, 2019, pp. 171105 – 171113.

[39] D. P. Kingma, J.L. Ba, "A Method for Stochastic Optimization", in International Conference for Learning Representations, 2015, arXiv:1412.6980v.

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 24-30
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

An Autoencoder based Emotional Stress State Detection Approach Using Electroencephalography Signals

Jia Uddin[1*]

[1]. AI and Big Data Department, Endicott College, Woosong University, Daejeon, South Korea

Received: 17 Nov 2021/ Revised: 04 Mar 2022/ Accepted: 27 Apr 2022

Abstract

Identifying hazards from human error is critical for industrial safety since dangerous and reckless industrial worker actions, as well as a lack of measures, are directly accountable for human-caused problems. Lack of sleep, poor nutrition, physical deformities, and weariness are some of the key factors that contribute to these risky and reckless behaviors that might put a person in a perilous scenario. This scenario causes discomfort, worry, despair, cardiovascular disease, a rapid heart rate, and a slew of other undesirable outcomes. As a result, it would be advantageous to recognize people's mental states in the future in order to provide better care for them. Researchers have been studying electroencephalogram (EEG) signals to determine a person's stress level at work in recent years. A full feature analysis from domains is necessary to develop a successful machine learning model using electroencephalogram (EEG) inputs. By analyzing EEG data, a time-frequency based hybrid bag of features is designed in this research to determine human stress dependent on their sex. This collection of characteristics includes features from two types of assessments: time-domain statistical analysis and frequency-domain wavelet-based feature assessment. The suggested two layered autoencoder based neural networks (AENN) are then used to identify the stress level using a hybrid bag of features. The experiment uses the DEAP dataset, which is freely available. The proposed method has a male accuracy of 77.09% and a female accuracy of 80.93%.

Keywords: EEG Signals; Emotion Analysis; Stress Analysis; Autoencoder; Machine Learning.

1- Introduction

For engineering wellbeing, detecting consequences from human error is essential because unsafe and irresponsible manners of employees involved in manufacturing are clearly accountable for human-caused troubles. Among several key factors of these dangerous and irresponsible activities, lack of proper sleep leads a person to an extreme stressful situation. Stress initiates irritation, fear, sadness, vascular illness and numerous additional injurious effects [1], [2]. Numerous forms of brain signals, i.e., functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS), Electrocorticography (ECoG), and electroencephalogram (EEG), are utilized for evaluating emotional conditions of individual [3]. Among all of these forms of data, EEG can be assessed non-intrusively [4]. The principal objective of this research is to categorize the emotional situation of an individual based on the sex by evaluating pre-processed freely accessible EEG signals.

Several surveys have exhibited relationships between EEG signals and several emotional situations [5–10]. In [5], an EEG-based assessment on the frontal channel with support vector machine (SVM) is designed. In [9], an in-depth analysis of power spectral density (PSD) is proposed to classify the emotional state by SVM. Among these researches, the common attribute is to consider all the features for classifier.

In this research, an EEG signal-driven emotional state classification method is established to evaluate whether a person is experiencing stress. By evaluating the signal, a hybrid feature bag is designed to create a dynamic and robust feature list. This process is divided into two parts: (1) statistical analysis from the time domain, and (2) wavelet-based feature assessment from the frequency domain. In the EEG signals, for the presence of the artifacts [11], it is hard to find the absolute feature information. This study examined pre-processed signals from the Database for Emotion Analysis Using Physiological Signals (DEAP) dataset [12]. The time

✉ **Jia Uddin**
jia.uddin@wsu.ac.uk

domain features are defined in detail in Section 3. As the EEG signal has five indistinguishable bandwidths [13], [14], wavelet decomposition is studied to determine the frequency domain features, which are depicted in Section III. From these designed bag of hybrid features, instead of providing all of them to the classifier, some built-in feature reduction mechanism embedded classifier is used to utilize only the most significant features for final classification. Deep networks can obtain extremely characteristic features via their multi-layered model architectures. Moreover, they keep only the most representative information in each layer to reduce the dimensionality and also to improve the classification performance by selecting only the most intrinsic feature information [15]. In this research, a three layered autoencoder based neural network (AENN) is proposed for different emotional state classification. To prove the strength of the suggested technique, few comparisons are made with the approaches discussed into [5], and [9]. This paper's primary contributions can be summarized as follows:

(1) Statistical analysis in the time domain and wavelet-based feature assessment in the frequency domain combine to create a hybrid bag of features.

(2) A two layered AENN is proposed to learn and utilize only the important features through the embedded feed-forward feature selection architecture to improve the final classification accuracy.

The remainder of this paper will be organized as follows. The DEAP dataset is readily available, and the recommended approach is explained in Section 2. Section 3 contains the data agreement, evaluation of the experimental findings, and discussion, and Section 4 concludes the paper.

2- Proposed Method

In Fig. 1, a block diagram proposed approach is provided. The proposed approach is divided into four sections: (1) pre-processed data gathering [12], (2) data arrangement, (3) creation of hybrid bag of features, and (4) AENN-based classification.

The data is first down sampled to 128 Hz, and then the artifacts are eliminated from the data, as seen in this diagram. The current study's annotation was completed after filtering the data using bandpass frequency and common segmentation. The statistical features from the time domain and wavelet-based frequency domain are then examined and retrieved from each class sample. Finally, an Autoencoder-based Neural Network is presented for classification.

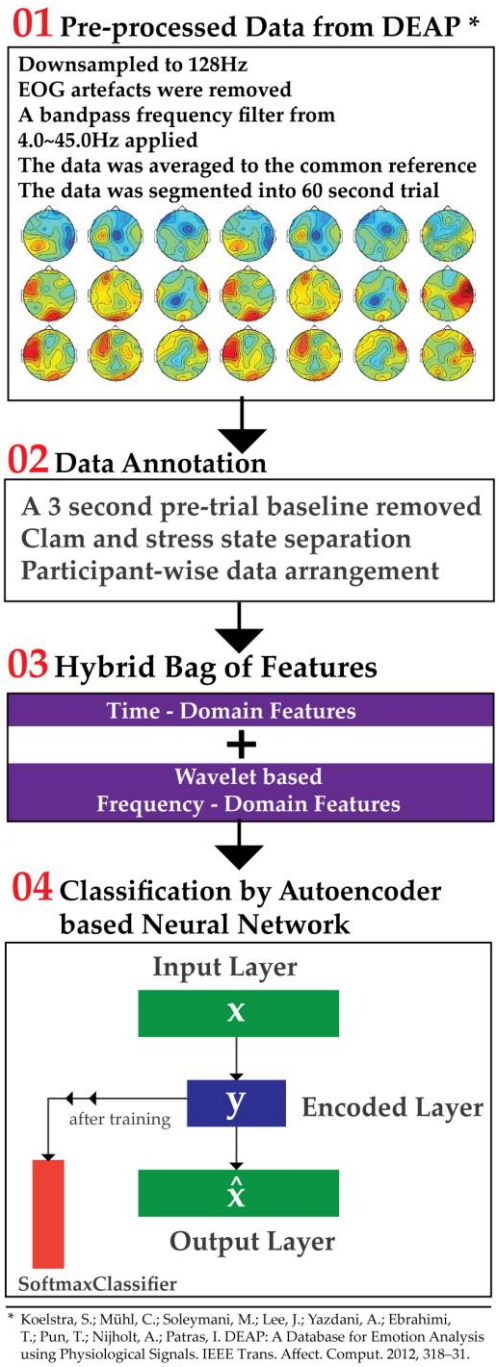


Fig. 1 Complete pipeline of the proposed approach.

2-1- Dataset Details

For this study, pre-processed EEG signals from the DEAP dataset [12]. This dataset incorporates emotional responses stimulated by music videos. The information related to the considered dataset is listed into Table 1. The placement of attached sensors for collecting the EEG data, with the real experimental setup is depicted into Fig. 2.

Table 1: Details of the considered dataset

Area	Information
Number of participants	32
Age range	19 – 37 years
Number of male participants	17
Number of female participants	15
Recorded time for EEG	1 minute 3 seconds
Pre-train baseline	3 seconds
Final considered signal length	1 minute
Electrode placement system	10 – 20 system of electrode

Each contributor viewed 40 music videos. The dataset retains pre-processed data that down sample the recorded signals to 128 Hz. In this research, the pre-processed EEG signals are considered. For this experimentation, the dataset has to be annotated. From the analyzed valence and arousal level from the recorded EEG response calm and stress states have been specified by Equation (1) and (2) [8], [16].

$$calm = (4 < valence < 6) \cap (arousal < 4) \quad (1)$$

$$stress = (valence < 3) \cap (arousal > 5) \quad (2)$$

Seven participants do not indicate any mental state of calm or stress after separating the data into two categories. The dataset is ready for the categorization mission, which is represented in Section 3, with the remaining over 25 contributors.

2-2- Hybrid Bag of Features

From time domain, the obtained numerical features are root mean square (F1), kurtosis (F2), skewness (F3), shape factor (F4), and impulse factor (F5). In addition to these, the mobility (F6) of the signal is also counted, which includes the information of the frequency spectrum [11], [17], [18]. The statistical details of these 6 features are described into Table 2.

EEG signals are distributed into 5 frequency bands, i.e., delta, theta, alpha, beta, and gamma [13], [14]. Here, as the sampling frequency of considered dataset is 128 Hz, level 5 wavelet decomposition facilitates to accumulate these 5 frequency bands. From the analyzed wavelet coefficients, wavelet energy (F7) and standard deviation (F8) is considered.

2-3- Classification by Autoencoder based Neural Network (AENN)

Autoencoder is mainly an unsupervised algorithm which learns the representation of the data by minimizing the reconstruction error from the layered architecture. It takes an input value x and then by using a function f in encoded the input value as y. Then, that encoded value turns into an output value $\hat{x}$, which is identical to the input. The main goal of autoencoder is to make the output value very similar to the input value by minimizing the reconstruction error. When it finally can be able to make the best reconstruction output, then the encoded value y from the encoded layer, learns the best data representation. In other words, it recreates the input from the encoded output appear in the encoded layer. Encoded layer produces a brand-new bag of features which is a mixture of the initial features. The encoded layer can be expressed by Equation (3):

$$k = g(wx + b) \quad (3)$$

Here, x is the input with the dimension of d, and then the encoded layer maps the input data to encoded latent variable k, where dimension reduced to d_k. w is weight and b is the bias here.

In this work, the same mechanism is deployed [15], [19]. The encoded layer latent feature representation is then passed to the SoftMax classifier for final classification. In the proposed AENN, two layers are used to learn the latent feature space in unsupervised way. The main architecture of the proposed AENN is illustrated into Fig. 3.

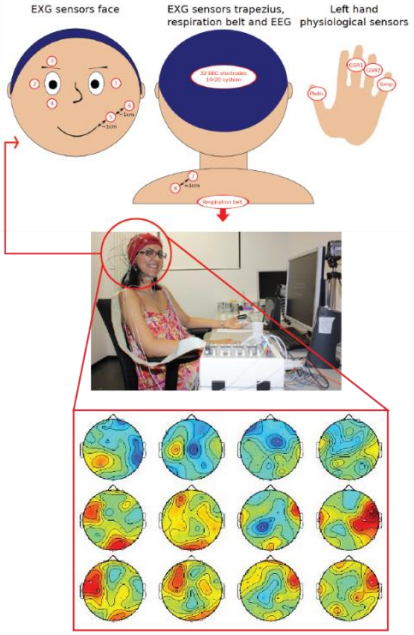


Fig. 2 Real experimental testbed with subject for collecting EEG data.

Table 2: Statistical explanation of the measured time-domain features

Feature	Equation	Feature	Equation	Feature	Equation
F1	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} X_i^2}$	**F2**	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\overline{X}}{\sigma}\right)^4$	**F3**	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\overline{X}}{\sigma}\right)^3$
F4	$\frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\overline{X}}{\sigma}\right)^3}{\frac{1}{N}\sum_{i=1}^{N}\lvert X_i\rvert}$	**F5**	$\frac{max(\lvert X\rvert)}{\frac{1}{N}\sum_{i=1}^{N}\lvert X_i\rvert}$	**F6**	$\frac{mobility(X')}{mobility(X)}$
Here, x is the time-domain raw signal. N is the total number of samples.					

Table 3: Particulars of the considered dataset

Dataset (Sex-based)	Sub-set	Participant ID	An Experimental ID System That Reflects the Unique State	
			Calm	Stress
Male	**1**	**1**	9,14	17,32,34,35,36,37
	2	**5**	13,29	23,30,37
	3	**12**	16,17,28	25,29,32,33,35,36,37,38
	4	**16**	6,12,16,21,36	1,15,17,24,26,27,34
	5	**18**	22,26,34	30
	6	**19**	15,26,27	29,38
	7	**20**	16,26,27,28,40	23,25,29
	8	**21**	3,21,26,34,35	20,22,24
	9	**26**	30	34
	10	**27**	5,15,19,26,27,28,33,40	27
	11	**28**	15,22,24,25	35,38
	12	**29**	15,17	30,31,33,35
Female	**1**	**2**	5,7,10,22,24,36	29,30,32,37,38
	2	**4**	2,6,18	24,28,32
	3	**8**	10,37,39	31,36
	4	**10**	15,17,20,22,26,27,28	21,30,35,36,37,38,39
	5	**11**	2,12,16,19,25,26,28,40	27,35,37,38,39
	6	**13**	12,15,16	7,21,23,31,34,35,36,37,38,39
	7	**14**	22,27	10,21,23,24,29,30,32,34,35,36,38
	8	**15**	7,16,22,26	24,25,30,38
	9	**22**	1,6,12,15,16,28	23,24,29,30,32,33,35,36,37,38,39
	10	**24**	33,40	21,23,24,30,31,38,39
	11	**25**	4,5,26,27,28,34	2,10,23,29,31,32,33,37,38,39
	12	**31**	17,22,24,27,28,29	23,32,34,37,38,39
	13	**32**	2,6,15,26,33	24,30,37

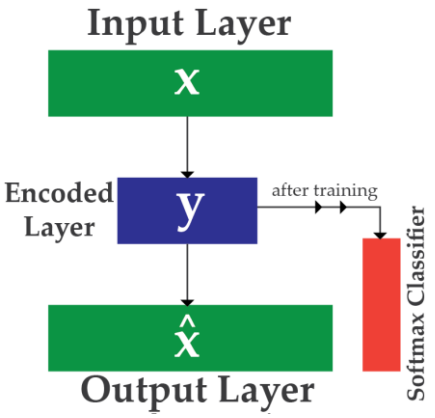


Fig. 3 AENN model architecture.

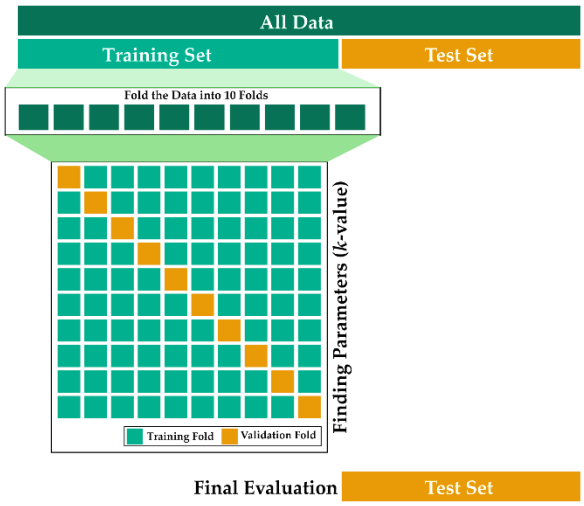


Fig. 4 10-fold cross validation [10].

3- Experimental Result Analysis

3-1- Arrangement of Dataset

Among 32 contributors [11], contributor numbers 3, 6, 7, 9, 17, 23, and 30 did not signify the attributes of considered emotional states. So, the left over 25 contributors are counted for the final dataset. 32 EEG channels are considered for considering the data with a 128Hz sampling frequency. The particulars of the considered dataset are given in Table 3. The Table reveals that each sub-dataset has a distinctive experimental ID, and for each ID, different music videos are reliable for calm and stress state. Therefore, for sex-based state analysis according to the identity of sex; two main datasets are considered for final classification, i.e., dataset male (consisting of 12 sub-sets, merged together to form male dataset as a whole), and dataset female (consisting of 13 sub-sets, merged together to form male dataset as a whole).

3-2- Performance Analysis of AENN

Each dataset is split into two parts: training and testing, with a 70/30 split between the two. Equation (4) establishes the class-dependent accuracy.

$$classdependent\_accuracy = \frac{true_{positive}}{true_{positive} + false_{positive}} \quad (4)$$

To ascertain the final average accuracy, Equation (5) is utilized.

$$average_{accuracy} = \frac{true_{posiotive} + true_{negative}}{total\ number\ of\ samples} \quad (5)$$

Using 10-fold cross-validation, the ultimate accuracy is calculated. K-fold [10] cross-validation, as illustrated in Figure 4, involves randomly dividing the training set into ten groups, or folds. The ultimate accuracy is calculated using 10-fold cross-validation. The training dataset is then divided into two parts: P for training and Q for validation. Training folds P are used to create the model, while validation fold Q is used to validate it. The AENN's parameters are tuned using the validation fold B. Every iteration (10 times), the validation fold is rotated, and the remaining data is utilized to train the AENN. The specifics of the final accuracy after 10-fold cross-validation are provided in Table 4.

Table 4: Classification accuracy of the proposed method

Dataset	Class-wise Accuracy (%)		Average Accuracy (%)
	Calm	Stress	
Male	78.25	75.93	77.09
Female	79.42	82.44	80.93
Average (%)			79.01

The feature embeddings extracted from the AENN encoded layer (2 features values are extracted from encoded layer) is displayed into Fig. 5 for both of the datasets.

In addition to these, to establish the robustness of this approach, few comparisons are made with [5] , and [9].The details of this comparative analysis are depicted into Table 5.

Table 5: Relative evaluation of various methodologies

Methods	Average Accuracy (%)	Decrement from the Proposed Method (%)
[5]	68.23	10.78
[9]	72.44	6.57
Proposed	79.01	-

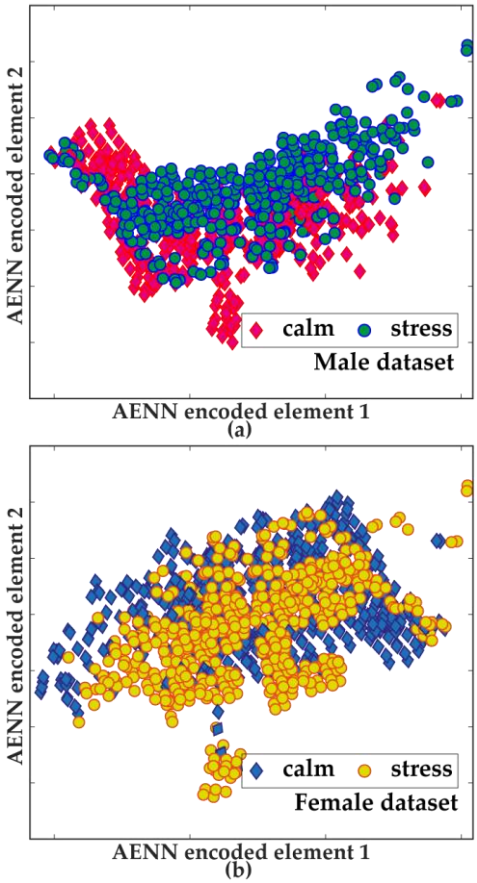


Fig. 5 AENN feature embedding from encoded layer for (a) male dataset, and (b) female dataset.

3-3- Discussion

From Table 4, it is visible that the female is performing better than the male dataset. From Table 3, it is visible that the female dataset contains 13 sub-sets where the male dataset contains 12 sub-sets. To get a better performance from deep networks, amount of dataset is a vital issue. Moreover, for female dataset case, stress state is giving higher accuracy than the calm state. From Table 3, it is clearly visible that the stress state data is larger than the calm state data for female dataset (61 cases for calm state and 84 cases for stress state).

From Fig. 5, for both of the dataset, the feature embeddings are overlapped. That indicates the separation of identical features are not good. The main underlying reasons behind this is little analytical. If the sub-sets are being analyzed from Table 3, for each sub-set, the experiment IDs of calm and stress state are different. In addition, sometimes it is imbalanced as well. So, in this dataset not any particular experiment ID is responsible for clam or stress state. It makes difficult to find out the most intrinsic class-wise information in a very accurate manner.

For comparative analysis, the proposed model outperformed the approach described in [5] by 10.78%, and the approach described in [9] by 6.57%.

4- Conclusions

This paper proposed a sex-based stress state classification method by analyzing EEG signals from brain. First it created a bag of features from the statistical analysis of time-domain and wavelet-based analysis of frequency domain. Therefore, the hybrid bag of features was forwarded to the proposed AENN to identify the stress state by feed-forward architecture based selective features. The proposed approach achieved an accuracy of 77.09% for male dataset and 80.93% for female dataset. In addition to that, it outperformed several existing methods related to this study by at least 6.57%.

Acknowledgments

This research is funded by Woosong University Academic Research in 2022.

References

[1] R. E. Bender, and L. B. Alloy, "Life stress and kindling in bipolar disorder: review of the evidence and integration with emerging biopsychosocial theories", Clinical psychology review, Vol. 31, No. 3, 2011, pp. 383–398.

[2] T. G. Pickering, "Mental stress as a causal factor in the development of hypertension and cardiovascular disease", Current hypertension reports, Vol. 3, No. 3, 2001, pp. 249–254.

[3] T. C. Major, and J. M. Conrad, "A survey of brain computer interfaces and their applications", in IEEE SOUTHEASTCON 2014, 2014, pp. 1–8.

[4] N. Elsayed, Z. S. Zaghloul, and M. Bayoumi, "Brain computer interface: EEG signal preprocessing issues and solutions", Int. J. Comput. Appl, Vol. 169, No. 3, 2017, pp. 12–16.

[5] R. E. Wheeler, R. J. Davidson, and A. J. Tomarken, "Frontal brain asymmetry and emotional reactivity: A biological substrate of affective style", Psychophysiology, Vol. 30, No. 1, 1993, pp. 82–89.

[6] A. C. Atencio et al., "Computing stress-related emotional state via frontal cortex asymmetry to be applied in passive-ssBCI", in 5th ISSNIP-IEEE Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), 2014, pp. 1–6.

[7] F.P. George et al. "Recognition of emotional states using EEG signals based on time-frequency analysis and SVM classifier", International Journal of Electrical and Computer Engineering, Vol. 9, No. 2, 2019, pp. 1012-1020.

[8] T. F. Bastos-filho, A. Ferreira, A.C. Atencio, S. Arjunan, and D. Kumar, "Evaluation of Feature Extraction Techniques in Emotional State Recognition", in 4th International conference on intelligent human computer interaction (IHCI), 2012, pp. 1-6.

[9] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Real-time EEG-based happiness detection system", The Scientific World Journal, 2013, pp. 1-12.

[10] M. J. Hasan, and J. M. Kim, "A hybrid feature pool-based emotional stress state detection algorithm using EEG signals", Brain sciences, 2019, Vol. 9, No. 12, pp. 1-15.

[11] D. Shon, K. Im, J. Park, D. Lim, B. Jang, and J. Kim, "Emotional Stress State Detection Using Genetic Algorithm-Based Feature Selection on EEG Signals", Int. J. of environmental research and public health, 2018, Vol. 15, No. 11, pp. 1-11.

[12] S. Koelstra, C. Muhl, M. Soleymani, J.S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals", IEEE transactions on affective computing, 2011, Vol. 3, No. 1, pp.18-31.

[13] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen, "EEG-based emotion recognition in music listening", IEEE Transactions on Biomedical Engineering, 2010, Vol 57, No. 7, pp. 1798-1806.

[14] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG", IEEE Transactions on Affective Computing, 2014, Vol. 5, No. 3, pp. 327-39.

[15] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A Hybrid Feature Model and Deep-Learning-Based Bearing Fault Diagnosis", Sensors, 2017, Vol. 17, No. 12, pp. 1-16.

[16] J. A. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology, 1980, Vol. 39, No. 6, pp. 1161–1178.

[17] S.H. Oh, Y.R. Lee, and H.N. Kim, "A novel EEG feature extraction method using Hjorth parameter", Int. Journal of Electronics and Electrical Engineering, 2014, Vol. 2, No. 2, pp. 106-110.

[18] P.C. Petrantonakis, and L.J. Hadjileontiadis, "Emotion Recognition from EEG Using Higher Order Crossings", IEEE Transactions on Information Technology in Biomedicine, 2010, Vol. 14, No. 2, pp. 186–197.

[19] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis", Mechanical Systems and Signal Processing, 2017, Vol. 95, pp. 187-204.

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 31-40
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Providing a New Smart Camera Architecture for Intrusion Detection in Wireless Visual Sensor Network

Meisam Sharifi Sani[1*], Amid Khatibi Bardsiri[2]

[1]. Department of Computer Engineering, Tehran Branch, Islamic Azad University, Tehran, Iran
[2]. Department of Computer Engineering, Bardsir Branch, Islamic Azad University, Kerman, Iran

Received: 20 Jan 2021/ Revised: 04 Jan 2022/ Accepted: 27 Feb 2022

Abstract

The wireless Visual sensor network is a highly functional domain of high-potential network generations in unpredictable and dynamic environments that have been deployed from a large number of uniform or non-uniform groups within the desired area, cause the realization of large regulatory applications from the military and industrial domain to hospital and environment. Therefore, security is one of the most important challenges in these networks. In this research, a new method of routing smart cameras with the help of cloud computing technology has been provided. The framework in the cloud computing management layer increases security, routing, inter interaction, and other features required by wireless sensor networks. Systematic attacks are simulated by a series of standard data collected at the CTU University related to the Czech Republic with RapidMiner software. Finally, the accuracy of detection of attacks and error rates with the suggested NN-SVM algorithm, which is a combination of vector machines and neural networks, is provided in the smart cameras based on the visual wireless sensor networks in MATLAB software. The results show that different components of the proposed architecture meet the quality characteristics of visual wireless sensor networks. Detection of attacks in this method is in the range of 99.24% and 99.35% in the worst and best conditions, respectively.

Keywords: Intrusion Detection; Smart Cameras; Security; Visual Sensor Network; Cloud Computing.

1- Introduction

Sensor networks consist of several nodes, which have computational, communicational, and sensing capabilities. Usually physically small, and it is limited in the processing power, memory capacity, and power supply [1]. These limitations provide some issues, which many research efforts in this area are stemmed from. The dense extension of sensor nodes should allow the network to coordinate with unpredictable environments. A type of wireless sensor network is a visual or video sensor network [2]. since embedding inexpensive cameras with (high or low) resolution features in wireless sensors, it is possible to receive the visual data from the environment, making a new focal point for the network applications [3]. The unique feature of the sensors equipped with video cameras is imaging a target or a part of a region which are not necessarily near the camera. That is, the cameras are capable of capturing objects in extra in their line of sights [4]. Accordingly, in these networks, Closed Circuit Television cameras are used as a node or a group of nodes for different applications. Today, one of the systems used for controlling and monitoring life and workplaces as well as providing more security and relief is a Closed-Circuit Television camera. These systems are referred to as either imaging control systems or closed-circuit video equipment. According to the setting applied to the cameras and other equipment, the systems are highly applicable in different climatic conditions at days and nights [5]. In the wireless sensor networks, each sensor node acquires a specific standpoint [6]. The sensor standpoint of the environment is limited by the range and precision, possibly covering a restricted physical area. Therefore, area coverage also can be considered as an important design parameter in wireless sensor networks. Since sensor nodes may produce considerable redundant data, the similarly produced data packets from numerous nodes can be integrated to reduce the number of transmissions [7]. Data integration is a combination of data from different sources based on a given integration function. This method has been used in some routing protocols to increase energy efficiency and to optimize the data transfer. Signal

✉ **Meisam Sharifi Sani**
Meisam.sharifi.s@gmail.com

processing methods could also be used to integrate the data [8]. This approach is also referred to as data combining, in which the node can produce more precise output signals by using different methods to combine input signals and to remove the noise of the signals [9]. In other words, Closed Circuit Television cameras are small devices with onboard predefined sensors, capable of image processing, which are publically accepted [10][11]. Some of the most important tasks of the Closed Circuit Television cameras in the wireless networks are full environmental control and monitoring, providing a security basis, traffic control, tracking, etc. [12]. Many researchers have managed and controlled numerous processes in the networks by aiding the cameras [8][13]. On the other hand, considering limitations before visual sensor networks with smart Closed Circuit Television cameras, there are some factors such as routing, security, interactivity, customization, etc. [14]. Intrusion detection systems are one of the possible solutions to solve security problems in wireless sensor networks. An Intrusion Detection System is also referred to as a second line of defense, which is used for intrusion detection only; that is, Intrusion Detection System can detect attacks but cannot prevent or respond. Once the attack is detected, the IDSs raise an alarm to inform the controller to take action. There are two important classes of IDSs. One is rule-based Intrusion Detection System and the other is anomaly-based Intrusion Detection System. Rule-based Intrusion Detection System is also known as signature-based Intrusion Detection System which is used to detect intrusions with the help of built-in signatures. Rule-based Intrusion Detection System can detect well-known attacks with great accuracy, but it is unable to detect new attacks for which the signatures are not present in intrusion database. Anomaly based IDSs detect intrusion by matching traffic patterns or resource utilizations. Although anomaly based IDSs have the ability to detect both well-known and new attacks, they have more false positive and false negative alarms. Some IDSs operate in specific scenarios or with particular routing protocols [15].

2- Problem Expression

Security and routing in sensor networks are very challenging due to the inherent characteristics of these types of networks, distinguishing them from other networks such as ad-hoc mobile networks or cellular networks that can be divided into 9 features, the most important differences of this network with other wireless networks. The first characteristic is that due to a large number of sensor nodes, it is not possible to build an addressing scheme. Therefore, the traditional IP-based protocols are not applicable for wireless sensor networks, because the ID maintenance overhead is very high. In the wireless sensor networks, sometimes receiving data is more important than understanding the data IDs which should be sent [16][17]. Secondly, sensing nodes located by the ad-hoc method should be self-organizing because the nodes ad-hoc locating requires systems to communicate and manage the resultant node distribution, especially the sensor network should be operated unsupervised [13]. Thirdly, in contrast to the conventional communication networks, almost all of the wireless network applications require transmitting the sensed data from numerous sources to a specific basic station. However, this characteristic does not prevent data flow in the other forms (for example, sending data in a multi-segment form or point to point) [18]. Fourthly, the sensor nodes are severely limited in energy, processing, and storage. Therefore, they need stringent resource management [19][20]. Fifthly, in the most applicable scenarios, the sensor network nodes except some of the mobile ones are usually fixed after locating. Nevertheless, in other traditional networks, the nodes are free to move, changing the topology frequently and unpredictably. In some usages, some sensor nodes can move and change their location (although with low mobility) [17]. Sixthly, sensor networks are pragmatic (i.e., the design requirements of the sensor network are changed based on the application of interest). For example, the challenging observation problem, which must be performed precisely with low delay, is different from periodic climate monitoring [5]. Seventhly, being informed about the sensor node's situation is highly important because data gathering usually relies on the location. This conditional awareness can be met by GPS devices although other methods independent from GPS are also practical for locating the problem in sensor networks [13]. Eighthly, data gathered by many sensor nodes in the wireless sensor network are typically related to a single phenomenon. Hence, the possibility of redundancy in the data is high. This redundancy should be extracted by the routing protocols to increase the usage of energy and bandwidth [13]. Finally, the wireless sensor networks are data-based because the data are requested based on the given features (i.e. feature-based addressing). A feature-based address consists of a query set of feature-value pairs [18]. Nowadays, researches on sensor network security field mainly focus on the following aspects:

1. Various attack resources such as networks, files, system logs, and processes cannot be used in wireless sensor networks, and we need to consider the feature information which can be applied to the wireless sensor network intrusion detection.
2. There are many new attacks in wireless sensor networks, which are different from traditional networks. How to improve the ability of the intrusion detection system to detect unknown attacks and select appropriate algorithms is a problem to be solved. Some algorithms are suitable for detecting known attacks, while others are suitable for detecting unknown attacks. Some algorithms are suitable

for a flat surface network structure, and some algorithms are suitable for a hierarchical network structure. We should select or design the appropriate algorithm according to the requirements of the network.

3. Wireless sensor networks have limited resources, including storage space, computing power, bandwidth, and energy. Limited storage space means that it is impossible to store large amounts of system logs on sensor nodes. The intrusion detection system based on knowledge is required to store large amounts of defined intrusion patterns. The system detects intrusion using pattern matching, and invasion behavior characteristics need to be stored in libraries. With the increase in invasion types, the scale of the feature library will also increase. Limited computing power means that the node is not suitable for running the intrusion detection algorithm which requires a lot of computation [21].

According to the mention of these features, it is known that routing is different in wireless visual networks from other wireless networks. Therefore, sending accurate information and securing routing in these networks can be useful. In this paper, we use an intrusion detection method based on ensemble learning algorithms in smart cameras based on wireless sensor networks. In this method, we use NN-SVM which is a combination of Support Vector Machine and Neural Networks.

3- Related Work

Various platforms have been proposed for wireless visual sensor networks. In this section, we study some of the most important cases proposed by different authors.

One study proposed a prototype of a cloud-based video recorder system under the Infrastructure as a Service (IaaS) abstraction layer in the cloud computing domain. This framework integrates a distributed file system to achieve a scalable, reliable, and virtualized architecture. In implementation design, Hadoop HDFS and HBase are integrated to reach scalability demand. This system provides scalable video recording, backup, and monitoring features. This framework integrates a distributed file system to achieve a scalable, reliable, and virtualized architecture. The paper shares a good experience while design an unlimited resource video recording system. In this method, the consumer bandwidth of the same quality of images as well as storage space in large networks with other methods has been used. They use open-source software, FFmpeg to transcending video resolution and format dynamically. The main problem with this software is that it is a CPU-compressed program, which can cause system bottlenecks if more people access the video from the same server [22].

Another study suggested a character order-preserving (COP)-transformation technique that allows the secure protection of video meta-data. The proposed technique has the merits of preventing the recovery of original meta information through meta transformation and allowing direct queries on the data transformed, increasing significantly both security and efficiency in the video meta-data processing, where the proposed technique was implemented for a real-world environment application, and its performance was measured. The method has the merit of increasing video meta-data efficiency significantly by allowing database query to take place in the same way as that adopted for plain-text files, without leaving the plain-text files exposed. The proposed mechanism accommodates match query, range search, and aggregation just as plain-text data would allow; furthermore, it allows the use of database indexes as they are and thus ensures processing efficiency. The proposed method characteristically divides the data into multiple chunks during encryption. A video meta-data query can therefore be executed based on the chunk ID that is assigned to the part of the video file that corresponds to the segment of the video footage. The proposed mechanism carries out decryption by avoiding the decoding of the whole Closed Circuit Television video data and instead obtains only the partial video data that satisfies the chunk-based search parameters. Efficiency, therefore, is ensured by the new method in terms of the data decoding performance [23].

In this research, the security method of cloud-based wireless IP cameras has been investigated, and the types of open and non-commercial text tools have been used in their research. They chose NetCam because of its ease of setup and use as well as low cost, which makes it a potential item for home-networks. They have captured and investigated the traffic generated by this device, both in idle and during live streaming to a mobile device over the cloud. The security of this device has been investigated from several research areas including secure multimedia, network security, and cloud security. Authors have checked the traffic generated by the low-end, easy-to-setup, off-the-shelf wireless IP camera for the average home user. This method examines the precautions taken by IP camera manufacturers and evaluates the mechanisms of access control in place. Many security issues and privacy in the use of these devices ranging from minor to severe issues have been identified. The results are achieved, suggesting that recently popular and easy to set up, easy-to-use effective cloud-connected wireless IP cameras should not run security and privacy [24].

In this work, they have proposed a video surveillance architecture based on the idea of cloud computing. By using this approach, it is possible to provide the video surveillance as a service. This architecture is the possibility of having a portable and scalable system that can be used in various scenarios. This system is based on the middleware FIWARE and has been implemented in a real scenario. In addition, as a result of video analysis it is

possible to obtain explanations of what is happening in a monitored area, and then take the appropriate action based on that interpretation. On the other hand, with this approach they are also able to add different kind of sensors besides cameras, this method provided a management panel digital devices as in a IoT framework [25].

4- System Model

The security requirements in a routing algorithm should be able to establish a route appropriately and to maintain it. This means that it should not allow the invader nodes to prevent the construction or proper maintenance of the route. If an algorithm meets the following points, it can be called a secure algorithm [26].

1. The routing signal could not be forged.
2. The manipulated signals could not be injected into the network.
3. The routing messages do not be changed during transfer except in the ordinary procedure of the protocol.
4. Routing loops do not be created during invading actions.
5. The invader nodes do not change the shortest routes.
6. The impermissible nodes should be removed from the network. This is assumed considering that; network management has a role in initializing and distributing keys, etc.
7. The network manager to neither permissible nodes nor invader ones should not present the network topology because they can use it to destroy the network.

According to the abovementioned requirements, in this section, the proposed method is described by the flowchart. It is followed by detailing the steps in the flowchart. In this section, the total architecture is studied which can be defined to integrate the architecture of visual sensor network framework and Closed-Circuit Television cameras by cloud computation.

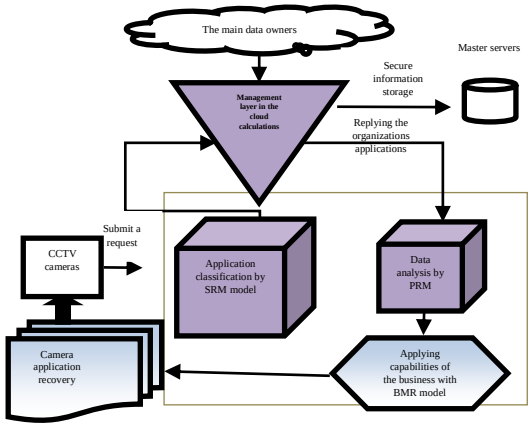


Fig. 1. The proposed reference model for locating the architecture of smart cameras in the cloud computation management layer [14]

As it can be seen in Fig 1, a smart camera routing in the wireless visual sensor network in the private environments is shown. These cameras outsource their data to the cloud computing environment, then this environment encodes the data and places them on its central and main servers. Each central server also consists of a series of hosts, providing the capability of data storage and management.

The sensor networks send their applications from the cloud computation in form of a query to the cloud computation management layer. This layer combined with some reference models in cloud computation send securely the received queries to the cloud computation. After recovering the data from the source or master servers, the recovered data are resent to the management layer in the cloud computation and are classified by the reference model [27], then are delivered to the smart cameras. In the next section, different parts of the proposed method are discussed and detailed.

4-1- Proposed Method

Considering the importance of cloud computation, in this section, the layer is studied and the location of the architecture of the smart camera is identified. One of the most important and applicable layers to locate the smart camera architecture in the wireless sensor network is the management layer in cloud computation. Layers in the architecture of the cloud computing reference Model in the proposed design.

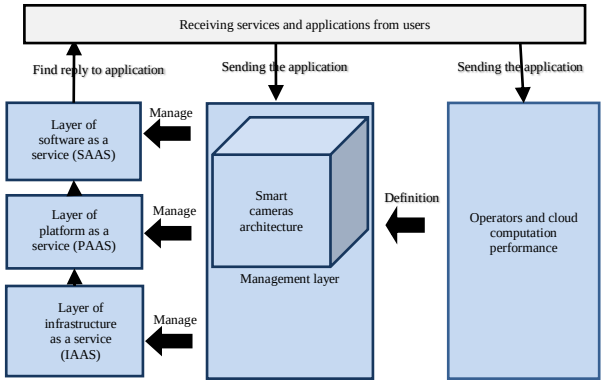


Fig. 2. The CCTV camera reference architecture model in the visual sensor networks and cloud computation.

Fig 2, the sent smart cameras routing to the cloud computation sends the reference application model to the management layer and operator layer. If any specific action is needed to be performed on the application, it is done in the operator layer, and the result is sent to the management layer. Considering the application type in the layer, in which the architecture of the smart camera is also located, the required management is applied according to the wireless visual sensor network, and the model of

interest is selected for the needed operations on the camera routing. Finally, delivering the client's response.

4-2- Proposed Method Pseudo Code

As earlier mentioned, the layer of infrastructure as a service has much more connected to the cloud computation management layer. According to the research need, it is assumed necessary to provide smart cameras locating steps in the cloud computation management layer in form of a pseudo code. In this section, according to the proposed innovation in the previous steps as well as following the provided flowchart and descriptions, the algorithm of interest is presented.

```
Function My_FEAF2.0_Cloud (Request R, String LN)
{
Input:
R: request of customer;
LN: Location Name of Customer;
Initialization:
C=CreateRequest (R, LN);
SPR_Model(C);//category Request
Master=SendRequest (R, LN) ;//send request to cloud and select master server
SD=Service_Delivery(R);//delivery request with Service Delivery Layer
SML=SendToManagementLayer (SD);
Process_Request_With_MLayer (SML);
SendToInfrastructureLayer (SML);
Tasks=Partitioning_Requests (SML);
Initialize Network Unit;
Initialize Storage Unit;// initialize Units of Infrastructure Layer
Initialize Compute Unit;
i=0;
While (Not End of Tasks)
{
CT=Compute (Tasks[i]);
Storeage_Cell[i]=StoreCTResult(CT);
i++;
}
SendToNextLayer (Storeage_Cell);
SendEndResultTo_DeliveryLayer (Result, R);
AR=Analyse_Result (Result, R);
SetBMR (AR);
Delivery_EndResultByCustomer (AR);
Output:
Result of Request;
}
```

ALGO 1. The proposed method algorithm

As it can be seen, the ALGO 1 is precisely based on the architecture and flowchart provided in the previous section, and the algorithm implementation steps are proportional to the steps provided in the previous section. In the input section of this pseudo-code, the application and user's name are received in the next step of the application's request function and the type of application will be checked after the initial examination is sent to the cloud management layer. The processing management layer is carried out and sent to the infrastructure layer in this layer. After that, it is calculated by the working computing unit, and the phase is sent to the storage unit. Finally, it shows the analyzed result.

5- Simulation and Experimental Results

In this section, the available cloud reference architectures and the proposed one are compared. It is noteworthy that the provided assessment is based on the comparison of the factors of architectures. Only some important measures are assessed in the comparison. Then, the proposed method in this paper is simulated using, RapidMiner software, and compared to other detection methods for security evaluation, detection of attacks in MATLAB software. The specifications related to the system by which the proposed method was implemented and the results were evaluated are shown in table 1.

Table 1. The system specifications for simulation and results assessment

Hardware/Software	Specifications
Operating system	Windows 7
Operating system type	32 bytes
RAM	4 GB- 3.06 usable
Processor	Intel processor- 7 (Core™) i7 CPU –) Q 720 @ 1.60GHz 1.60 GHz

5-1- Proposed Method Assessment According to the Security Features

In the internet-based environments, preserving users' security and privacy is of great importance. This is more important for the clients who provide their information to the cloud via smart cameras. In the proposed architecture that the combination of vector machines and neural networks algorithm, a simulation was performed by using Rapid Miner software to test and evaluate the data transfer security and providing services to the client through smart cameras. The obtained results are discussed in the following.

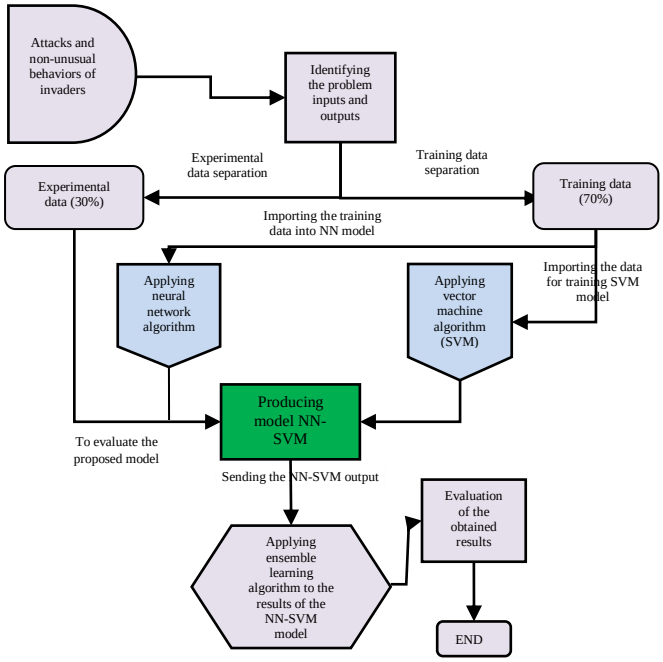


Fig. 3. The flowchart of applications security for CCTV cameras of the cloud computation by using the proposed NN-SVM method

Fig 3, the flowchart of applications and transfers security for smart cameras in visual networks in the proposed cloud computation architecture is shown.

The flowchart in Fig 3 shows the identifying procedure of the attacks to the smart cameras in visual networks according to the proposed method, which is combined from the support vector machine algorithm, and neural network as well as cumulative strategy. In this research, the proposed method is named NN-SVM. Considering the flowchart, first, the dataset related to the applied attacks on the smart cameras in visual networks, so-called the dataset of attacks to the system is imported to evaluate the proposed NN-SVM method. After that, the problem inputs as variables or features related to the dataset are identified. The feature indicating the attack type or application normality is also determined.

To evaluate the security and to model the proposed framework in this research, it was needed to use some datasets for testing the security. Therefore, there are not so many datasets to evaluate whether abnormal behaviors of the systems in the smart cameras exist in the visual networks. Studying basic used kinds of literature in this research, it was seen that most of the papers have exploited Malware Capture Facility Project (MCFP) dataset. This dataset was compiled at CTU University in the Czech Republic [28]. The files on each dataset are usually very large so they are stored in a server at the university. This dataset has various versions from which the newest one was used for evaluation. This is the most commonly used dataset for detecting systems abnormal behaviors and for evaluating the security of architecture in cloud computation. MCFP dataset includes 1048576 records with 39 features.

Out of 39 available features, 38 are defined as the system input and one is defined as the system output. By logging of the 39 features, 15 ones are selected as prominent features by the university. The first 14 features summarized in the following table indicate the features related to a system's abnormal behavior in form of an attack. The 15th feature as the last one is as the placed attack. Using the first 14 features, the last one (label) can be identified.

Due to the large size of the feature set and to facilitate the evaluations, 10 percent of the records were chosen randomly for the chosen set to include 10500 records with the same features. The dataset includes all of the attacks such as Botnet, DDOS, CVUT[1], etc. which are occurred for each system, organization, or individual's application. For correct evaluation, all of the attacks are considered as the same which 1654 records did not contain attacks, and 8855 records include attacks and system abnormal behaviors. Various available features in the dataset are summarized in Table 2.

Table 2. List of features and related characteristics

Feature name	Description	Type
Star Time	Start time	Continuous
Dur	Connection duration	Discrete
Proto	Protocol	Discrete
srcAddr	Source Address	Continuous
Sport	Source Port	Continuous
Dir	Directory	Discrete
DstAddr	Destination Address	Discrete
Dport	Destination Port	Continuous
State	State	Continuous
sTos	Source application service	Discrete
dTos	Destination offered service	Discrete
ToPackt	Number of packets	Continuous
TotByes	Number of bytes	Continuous
SrxBytes	Size of packet	Continuous
Label	Label (output)	Discrete

The last feature, which is considered as the attack type applied on the smart cameras in the visual networks, is divided into 5 categories, showing the applied attack types. The attacks are classified as follows:

Number 1 shows normal behaviors. Other numbers from 2-5 show abnormal behaviors, which are Botnet, attacks. Therefore, Botnet attacks are divided into 4 categories, which are known as Botnet type 1, Botnet type 2, Botnet type 3, and Botnet type 4.

All of the 15 features mentioned are used as the training model, and the last feature of the label is used to identify the type of attacks and abnormal behavior of the system. According to the procedure, the related data including these 15 features as the training samples are imported in the simulator. After performing and applying the proposed methods to the

1. It is among the most important attacks and abnormal behaviors applied by the hackers on the smart cameras in the visual networks.

related data, the training samples are introduced with no label. In this step, considering 14 features of interest, attacks and system abnormal behaviors are identified, and a label is attached, showing what the type of attack is.

Then, the imported data are separated into two training and experimental sections. The training data are used to produce the models related to the support vector machine (SVM) and neural network (NN) algorithms among the most commonly used algorithms to identify breaches. The experimental data are exploited to evaluate the precision and error of the proposed method. To this end, 70% and 30% of the data are considered as the training and experimental data, respectively. Now, the training data are applied to the NN and SVM models to produce the related model. Finally, the experimental data are applied to the produced model, and for each sample of the experimental data, it is identified that the sample is whether an attack or a normal application. The cumulative strategy is applied to the results and outputs of the NN and SVM models, performing the last identification finally.

Considering the above flowchart, the proposed method was implemented by RapidMiner software. The modeling related to the SVM algorithm is seen in fig 4.

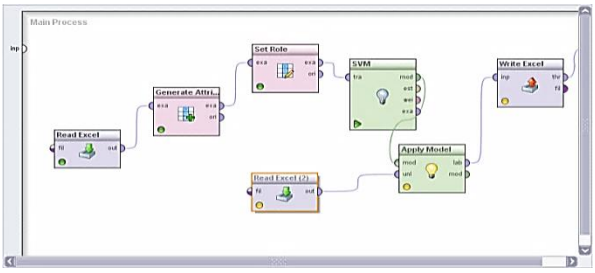


Fig. 4. SVM model for modeling and identifying attacks in the cloud computation

The training data of 70% are applied to the Generate Attribute control, determining the problem inputs. Then, the problem output is identified by the Set Role control. Finally, the SVM algorithm is performed on the training data, and the related model is produced as Apply Model. Then, Read Excel imports the experimental data into the bottom section; finally, it is applied to the produced model called Apply Model, generating the results. The models related to the NN algorithm are seen in Fig 5.

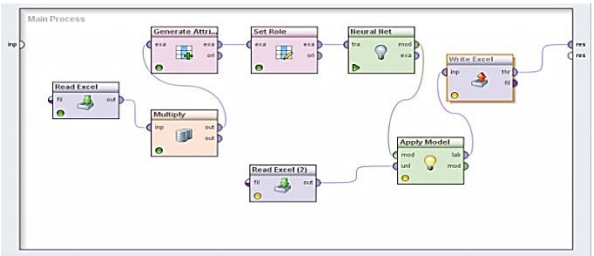


Fig. 5. The neural network model for generating the model and identifying attacks in the cloud computation

The assessment results for the proposed architecture security in the cloud computation by using the proposed NN-SVM method are as follows.

Table 3. Results and identification precision of the attacks by using the NN-SVM algorithm

Sample	Type of the main attack	Attack identification by SVM method	Attack identification by neural network	SVM error	NN error
1	1	1.659860319	1.046367543	0.65986	0.046368
2	1	1.604163382	1.01363806	0.604163	0.013638
3	1	1.586164829	1.005893113	0.586165	0.005893
4	1	1.576049401	1.003681723	0.576049	0.003682
5	1	1.570627125	1.002386739	0.570627	0.002387
6	2	2.683608079	1.298985549	0.683608	0.701014
7	2	2.676240453	1.993326485	0.67624	0.006674
8	2	2.6·2283178	1.298272655	0.682283	0.701727
9	3	4.618082618	3.222731178	1.618083	0.222731
10	3	2.997662624	2.820934385	0.002337	0.179066
11	3	3.166524806	2.914030606	0.166525	0.085969
12	5	6.006369533	4.742892198	1.00637	0.257108
13	5	5.995303345	4.826195301	0.995303	0.173805

Checking Table 3. the types of attacks to the smart cameras in visual networks were identified by using various methods. As an example, in row 13, the attack is of type 5 which is a storage database in the system. By applying the SVM algorithm to the related sample, the answer of 5.9 is obtained, which the final solution is acquired by rounding it up or down. By rounding 5.9, the answer is 6 to which the attack is not identified correctly; this is a negative point. However, the neural network identified it correctly as 4.8; that is a plus. Combining the two algorithms due to neural network algorithm has higher precision than the SVM, and because the answers are not similar, the solution of the model is chosen identifies better.

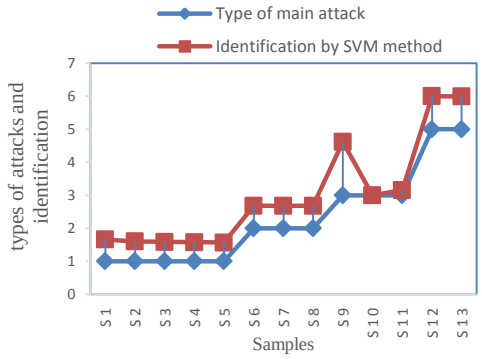


Fig. 6. Preserving the security of the smart cameras in visual networks in the proposed architecture by using a support vector machine algorithm

In fig 6, discrepancies of the SVM algorithm identification are shown for 13 samples of the applied attacks to the smart cameras in visual networks in MATLAB software. At each step you can see the number of detected attacks in the vertical axes, this number is compared to the type of

attacks stored in the database. Finally, the attack is detected by the SVM algorithm.

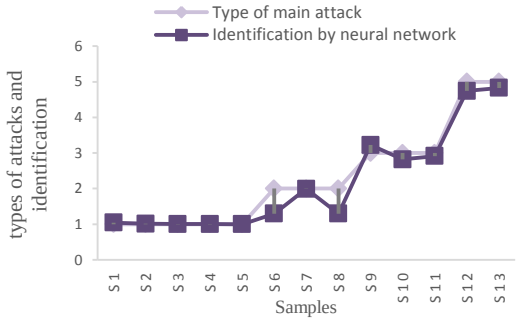


Fig. 7. Preserving the security of the smart cameras in visual networks in the proposed architecture by using a neural network algorithm

As it can be seen in fig 7, it is evident in figure 7 that attacks are identified by using the proposed algorithm in the architecture with a very low discrepancy, allowing breach detection to the proposed architecture. In the following figure, discrepancies of the neural network algorithm identification are shown for 13 samples of the applied attacks and in vertical axes has shown the number of obtained attacks in every step according to the proposed architecture security features in MATLAB software.

By viewing in Fig 8, attacks are identified by the proposed algorithm in the architecture with a very low discrepancy, allowing breach detection to the proposed architecture. It is noteworthy that the neural network algorithm has a much lower discrepancy compared to the support vector machine algorithm in attack identification.

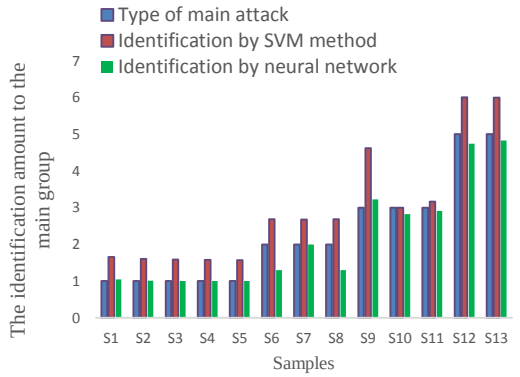


Fig. 8. Comparison of the precision of the proposed NN-SVM method for preserving the security of the proposed architecture

In Fig 8, the number of attacks identified using SVM and NN methods versus the main attacks to the smart cameras in visual networks is shown.

In Fig. 9. The error of attacks identification by using SVM and NN methods versus the main attacks are shown. In this way, we have been able to distinguish more accuracy by combining two vector machines and neural network algorithms, and the attacks that are sent towards smart cameras in visual wireless networks will be more accurate.

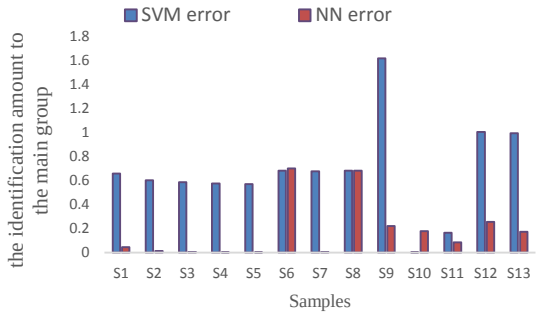


Fig. 9. Comparison of the error of the proposed NN-SVM method for identifying the attacks to the smart cameras in visual networks

Considering the calculation of the proposed method precision using Rapid Miner, the precisions were 99.24% and 99.35% in the best and worst conditions, respectively. Therefore, the proposed NN-SVM method for the architecture in this research preserves the data security for the smart cameras in visual networks with applicable precision. It defeats the attack when observing one of the proposed frameworks.

Table 4 shows the comparison of the results of the proposed method with other methods. In this research, the accuracy criterion is one of the important design criteria and we compare it with other methods. The table below shows the accuracy of the proposed method with other popular methods.

Table 4. Comparison of accuracy, call accuracy, and error detection to maintain security in the proposed method with other methods.

Detecting Attacks	**Proposed Method**	**Liner-SVM** [29]	**RBF** [29]
Accuracy	99.35%	98.19%	98.21%
Error	0.65%	1.81%	1.79%

In general, the two most important criteria for detecting attacks to maintain security in the proposed methods are accuracy and error detection, which, as shown in the table above, the proposed method in these two cases works better than the methods mentioned in the source [29]. As can be seen, the accuracy of the proposed method has improved by about 1.16% compared to the Liner-SVM method and has improved by about 1.14% compared to the RBF method.

6- Conclusion

A network of visual sensor networks is distributed from smart camera devices that can process and fuse images from a scene from different perspectives to some of the more useful forms of individual images. Visual sensor

networks are useful in applications that include area monitoring, tracking, and environmental monitoring. Visual sensor networks (VSNs) are receiving a lot of attention in research, and at the same time, commercial applications are starting to emerge. They use wireless communication interfaces to collaborate and jointly solve tasks such as tracking persons within the network. VSNs are expected to replace not only many traditional, closed-circuit surveillance systems but also to enable emerging applications in scenarios such as elderly care, home monitoring, or entertainment. The highly sensitive nature of images makes security and privacy in VSNs even more important than in most other sensor and data networks. In the current research, a novel framework is provided by aiding the cloud computing technology. Then, the detection of attacks on smart cameras in wireless sensor networks was analyzed with the proposed algorithm of NN-SVM. This method has been simulated with systematic attacks collected at the CTU University-related Czech Republic with RapidMiner software. Finally, the Accuracy of detection of attacks and the proposed method error was implemented in MATLAB software. The results show that the proposed algorithm has been able to provide security requirements for smart cameras in visual wireless networks.

References

[1] V. D. Kale, “Recent Research Trends in Cloud computing,” vol. 6, no. 2, pp. 406–409, 2013.

[2] A. Boukerche et al., “A new solution for the time-space localization problem in wireless sensor network using UAV,” in Proceedings of the third ACM international symposium on Design and analysis of intelligent vehicular networks and applications - DIVANet ’13, 2013, pp. 153–160, doi: 10.1145/2512921.2512937.

[3] Y. Charfi et al., “A pervasive smart camera network architecture applied for multi-camera object classification,” Proc. IEEE, vol. 96, no. 2, pp. 636–641, May 2009, doi: 10.1109/ICDSC.2009.5289377.

[4] B. Rinner, T. Winkler, M. Quaritsch, B. Rinner, W. Schriebl, and W. Wolf, The evolution from single to pervasive smart cameras Epigenetic regulation of stress induced drug tolerance View project VECTO-Vehicle Energy Consumption Calculation Tool View project THE EVOLUTION FROM SINGLE TO PERVASIVE SMART CAMERAS. 2008.

[5] P. Chen et al., “Citric: A low-bandwidth wireless camera network platform,” in 2008 2nd ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2008, 2008, doi: 10.1109/ICDSC.2008.4635675.

[6] P. Saastamoinen, S. Huttunen, V. Takala, M. Heikkilä, and J. Heikkilä, “Scallop: An open peer-to-peer framework for distributed sensor networks,” in 2008 2nd ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2008, 2008, doi: 10.1109/ICDSC.2008.4635712.

[7] “Mohajer, Amin, Maryam Bavaghar, Rashin Saboor, and Ali Payandeh. "Secure dominating set-based routing protocol in MANET: Using reputation." In 2013 10th International ISC Conference on Information Security and Cryptology (ISCISC), pp. 1-7. IEEE, 2013.”

[8] W. Dargie and C. Poellabauer, Fundamentals of Wireless Sensor Networks. 2010.

[9] P. Huang, L. Xiao, S. Soltani, M. W. Mutka, and N. Xi, “The evolution of MAC protocols in wireless sensor networks: A survey,” IEEE Commun. Surv. Tutorials, vol. 15, no. 1, pp. 101–120, 2013, doi: 10.1109/SURV.2012.040412.00105.

[10] K. Langendoen and N. Reijers, “Distributed localization in wireless sensor networks: a quantitative comparison,” Comput. Networks, vol. 43, no. 4, pp. 499–518, Nov. 2003, doi: 10.1016/S1389-1286(03)00356-6.

[11] F. Gherardi and L. Aquiloni, “Sexual selection in crayfish: A review,” Crustac. Monogr., vol. 15, no. August, pp. 213–223, 2011, doi: 10.1163/ej.9789004174252.i-354.145.

[12] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. 22, no. 1, pp. 1–10, Jan. 1976, doi: 10.1109/TIT.1976.1055508.

[13] S. Soro and W. Heinzelman, “A Survey of Visual Sensor Networks,” Adv. Multimed., vol. 2009, pp. 1–21, 2009, doi: 10.1155/2009/640386.

[14] M. Malathi, “Cloud computing concepts,” in 2011 3rd International Conference on Electronics Computer Technology, 2011, vol. 6, pp. 236–239, doi: 10.1109/ICECTECH.2011.5942089.

[15] N. A. Alrajeh, S. Khan, and B. Shams, “Intrusion Detection Systems in Wireless Sensor Networks : A Review,” vol. 2013, 2013, doi: 10.1155/2013/167575.

[16] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, “Wireless multimedia sensor networks: applications and testbeds,” in Proceedings of the IEEE, 2008, vol. 96, no. 10, pp. 1588–1605, doi: 10.1109/JPROC.2008.928756.

[17] M. Ahmadi and S. M. Jameii, “A Secure Routing Algorithm for Underwater Wireless Sensor Networks,” Int. J. Eng., vol. 31, no. 10, pp. 1659–1665, Oct. 2018, doi: 10.5829/ije.2018.31.10a.07.

[18] M. Quaritsch, B. Rinner, and B. Strobl, “Improved agent-oriented middleware for distributed smart cameras,” in 2007 1st ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC, 2007, pp. 297–304, doi: 10.1109/ICDSC.2007.4357537.

[19] A. Doblander, A. Zoufal, and B. Rinner, “A novel software framework for embedded multiprocessor smart cameras,” ACM Trans. Embed. Comput. Syst., vol. 8, no. 3, pp. 1–30, Apr. 2009, doi: 10.1145/1509288.1509296.

[20] S. R. Taghizadeh and S. Mohammadi, “LEBRP - A lightweight and energy balancing routing protocol for energy-constrained wireless ad hoc networks,” Int. J. Eng. Trans. A Basics, vol. 27, no. 1, pp. 33–38, 2014, doi: 10.5829/idosi.ije.2014.27.01a.05.

[21] R. Zhang, “Intrusion Detection in Wireless Sensor Networks with an Improved NSA Based on Space Division,” vol. 2019, no. 1, 2019.

[22] C.-F. Lin, S.-M. Yuan, M.-C. Leu, and C.-T. Tsai, “A Framework for Scalable Cloud Video Recorder System in Surveillance Environment,” in 2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted

Computing, 2012, pp. 655–660, doi: 10.1109/UIC-ATC.2012.72.

[23] J. Kim, N. Park, G. Kim, and S. Jin, "CCTV Video Processing Metadata Security Scheme Using Character Order Preserving-Transformation in the Emerging Multimedia," Electronics, vol. 8, no. 4, p. 412, Apr. 2019, doi: 10.3390/electronics8040412.

[24] A. Tekeoglu and A. S. Tosun, "Investigating Security and Privacy of a Cloud-Based Wireless IP Camera: NetCam," in 2015 24th International Conference on Computer Communication and Networks (ICCCN), 2015, vol. 2015-Octob, pp. 1–6, doi: 10.1109/ICCCN.2015.7288421.

[25] L. Valentín, S. A. Serrano, R. Oves García, A. Andrade, M. A. Palacios-Alonso, and L. Enrique Sucar, "A CLOUD-BASED ARCHITECTURE FOR SMART VIDEO SURVEILLANCE," ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. XLII-4/W3, no. 4W3, pp. 99–104, Sep. 2017, doi: 10.5194/isprs-archives-XLII-4-W3-99-2017.

[26] K. K. Basu, "Organisational culture and leadership in ERP implementation," Int. J. Strateg. Chang. Manag., vol. 6, no. 3/4, p. 292, 2015, doi: 10.1504/IJSCM.2015.075919.

[27] D. Bijwe, "International Journal of Computer Science and Mobile Computing Database in Cloud Computing-Database-as-a Service (DBaas) with its Challenges," Int. J. Comput. Sci. Mob. Comput., vol. 4, no. 2, pp. 73–79, 2015.

[28] M. Grill and J. Stiborek, "An Empirical Comparison of Botnet Detection Methods An Empirical Comparison of Botnet Detection Methods," no. September, 2014, doi: 10.1016/j.cose.2014.05.011.

[29] "Intrusion Detection System Based on Cost Based Support Vector Machine," 2016, doi: 10.1007/978-3-319-40415-8.

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 41-47
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

A Customized Web Spider for Why-QA Pairs Corpus Preparation

Manvi Breja[1*]

[1].Department of Computer Engineering, Gurugram University, Gurugram, Haryana, India

Received: 9 Feb 2021/ Revised: 20 Mar 2022/ Accepted: 27 Apr 2022

Abstract

Considering the growth of researches on improving the performance of non-factoid question answering system, there is a need of an open-domain non-factoid dataset. There are some datasets available for non-factoid and even how-type questions but no appropriate dataset available which comprises only open-domain why-type questions that can cover all range of questions format. Why-questions play a significant role and are usually asked in every domain. They are more complex and difficult to get automatically answered by the system as why-questions seek reasoning for the task involved. They are prevalent and asked in curiosity by real users and thus their answering depends on the users' need, knowledge, context and their experience. The paper develops a customized web crawler for gathering a set of why-questions from five popular question answering websites viz. Answers.com, Yahoo! Answers, Suzan Verberne's open-source dataset, Quora and Ask.com available on Web irrespective of any domain. Along with the questions, their category, document title and appropriate answer candidates are also maintained in the dataset. With this, distribution of why-questions according to their type and category are illustrated. To the best of our knowledge, it is the first large enough dataset of 2000 open-domain why-questions with their relevant answers that will further help in stimulating researches focusing to improve the performance of non-factoid type why-QAS.

Keywords: Non-Factoid Questions; Web Crawler; Latent Dirichlet Allocation; Topic Modeling; Natural Language Processing.

1- Introduction

Question Answering Systems (QASs) answer the users' questions asked in natural language. With the increase in popularity of information access and providing an ease to user with an appropriate answer
to question, there is a challenging issue of automatically answering non-factoid questions. In English language, there are two broad categories of questions (1) Factoid questions of type what, where, who, when, which are simple and answered in a single phrase or sentence and (2) Non-Factoid questions comprising why and how-type are complex involving explanations and detailed reasoning in their answers. A non-factoid question possesses different answers to satisfy users' curiosity of different knowledge backgrounds [39]. Promising results are achieved for answering factoid questions; however non-factoid question answering is a challenging task. They require advanced NLP techniques to resolve open issues such as (1) ambiguity, (2) variability and (3) redundancy. Ambiguity implies difficulty faced in interpreting the context of non-factoid questions, variability implies different possible forms of answers to non-factoid questions, and redundancy refers to retrieving unnecessary texts along with possible answers of non-factoid questions [38,40]. The prime requisite for implementing non-factoid question answering system is dataset preparation depending on the requirement of the task. Without an efficient dataset of questions and their answers, it is very difficult to contribute significantly to the research community.

There are various datasets available, some of which are restricted domain [2,4,14,27,28,29,30], some correspond to only how-type questions (for ex. Yahoo! Answers Manner Questions L4 and L5 [31]) and some comprising why-type questions (for ex. only 3% of wh-questions in MSN click data, only 4.7% of 17000 questions are why-type questions in Webclopedia data collection [16, 32]). The questions in the dataset are significantly less in number which is not appropriate for research in developing why-type question answering system. Therefore, this paper has tried to develop a dataset for only why-type questions and their answers that is not only significant for our research but will contribute to research community working in the field of why-QA.

✉ **Manvi Breja**
manvibreja91@gmail.com

The paper is organized into sections where Section 1 introduces the concept of question answering with factoid and non-factoid QAS. Section 2 discusses the motivation for research in why-question answering. Section 3 describes various existing factoid and non-factoid datasets. Section 4 discusses the process of crawling and scraping applied to prepare a dataset of why-QA pairs. Section 5 analyzes the distribution of topics and different forms of why-question. Section 6 highlights applications of why-QA dataset and finally section 7 concludes with future research directions.

2- Motivation for Research in Why-QAS

Why-type questions are complex and involve variations in their answers. Their answers seek reasoning and explanations about the entity involved in the question. These questions depict the curiosity about something which are usually asked in every fields of life. There are some type of questions which have a definite reasoning and some have multiple possible answers depending on the users' knowledge, context and their experience. These questions and their answers possess cause-effect relationships where cause part is depicted in the answers and its effect part is asked in the questions. A considerable performance is already achieved in factoid type question answering, however research in non-factoid question answering is still challenging. There are different categorizations of non-factoid questions provided by [33] viz. list, confirmation, causal and hypothetical. The paper carries out an approach in why-question answering that incorporates causal relations. They are difficult to get automatically answered as it is difficult to understand the accurate users' interpretations and thus require advanced techniques. There are limited number of open-source datasets available which comprises significant number of why-type questions. This motivates us to design a dataset having open-domain why-type questions and their answers.

3- Available Factoid and Non-Factoid Datasets

This section discusses the brief characteristics of the datasets available on web. It also highlights their merits and demerits which significantly motivates us to prepare a dataset for why-QA pairs.

Suzan Verberne in 2006 [1] released a first open-domain why-QA dataset on Web. It is an open-source dataset comprising 395 why-questions and 769 corresponding answers, out of which 166 questions were further paraphrased. In addition to the question, dataset also comprises its source document with relevant user-formulated answers. It is effective to be utilized by the researchers working on QAS but not large enough and needs to be expanded further. InsuranceQA in 2015 [2] comprises restricted questions on insurance domain. The questions are collected from insurance library website, comprising 16889 questions and 27413 answers. The questions are usually asked from real users and answers are constructed by domain experts. The dataset can be utilized for applying deep learning models on answer selection task. In 2018, a large dataset named WikiPassageQA [3] is developed which comprises 4165 open-domain non-factoid questions split into 0.8/0.1/0.1 resulting training set comprising 3332 questions, development set comprising 417 questions and testing set comprising 416 questions. With the questions, documentID, document name and answer passage are also stored in the dataset. It is effective to be utilized for answer passage retrieval from relevant documents by applying deep learning models. Further in 2018, another restricted domain why-QA dataset; PhotoshopQuiA [4] is released which comprises 2854 why-questions on Adobe Photoshop collected from five CQA websites, viz. Adobe Forums [6] Stack Overflow [5], Graphics Design [7], Super User [8] and Feedback Photoshop [9]. With each why-QA pair, dataset also includes question id, URL, title, date, questioner, his level, open or resolved state, full question text, HTML and for each answer: answer date, answerer, his level, answer votes, text, full answer HTML and a value indicating if answer is best or not. The dataset is effective to be utilized for understanding different characteristics of why-questions and further instigating them for recommendation systems and chatbots. Freebase QA in 2019 [10] is an open-domain QA developed by complementing trivia type QA pairs with Freebase triples, comprising 54K matches from 28,348 unique questions with 20,358, 3994, and 3996 training, development and evaluation sets respectively. The dataset comprises following entries version of dataset, set of unique questions in dataset comprising Question-ID, original question, processed question, semantic parse-id, topic entity in question, name of topic entity, Freebase MID of topic entity, path from topic entity to answer node in Freebase, Freebase MID of answer, answer string from original QA pair . It is effective enough to train machine learning models as QA pairs are matched with (subject, predicate, object) triples that also helps to understand the meaning of questions and search for correct answers in Freebase. It can be utilized for several applications like reading comprehension [11], natural language understanding and search. It is a complex knowledge-base dataset which makes invaluable for more advanced ML methods. ANTIQUE in 2020 [12] is a collection of 34,011 non-factoid QA pairs usually asked by users on Yahoo!Answers [13] where 2426 questions & 27.4k

judged answers are training set and 200 questions & 6.5k judged answers are testing sets. Out of 34,011 QA pairs, 36% of questions are why-questions majorly covering intent of questions asked on community QA sites. TutorialVisualQA in 2020 [14] comprises 6195 non-factoid QA pairs based on 76 tutorial videos. Dataset includes varied fields like questions, video_id, manually annotated answer fragments, answer_start and answer_end denoting indexes of beginning and ending answer sentences. This dataset can be helpful for implementing a model to understand answer boundary sentences which can further be utilized for other educational, instructional, cooking videos and many more.

The brief description and features of the available datasets comprising why-questions is discussed in Table 1.

Table 1: Brief of some existing datasets on why-questions and their answers

Datasets	Description
Suzan Verberne dataset (2006)	395 why-questions and 769 answers formulated by annotators, collecting text documents from Textline Global News (1989) [34] and The Guardian on CD-ROM (1992) [35]. Dataset is not large enough to be utilized for further research
InsuranceQA	Comprises 15867 non-factoid questions and 24981 unique answers related to insurance domain. Accurate number of why-questions in dataset is not mentioned. The dataset is restricted to domain thus can't be utilized for all research purposes
WikiPassageQA	Comprises total 4165 open-domain non-factoid questions and 244136 candidate passages as answers to these questions. Out of these only 14% are why-type questions which is definitely not cover all range of why-type questions and thus not significant for use in why-QAS.
PhotoshopQuiA	2854 why-QA pairs collected from 5 community QA websites. The dataset is restricted to Photoshop domain and mostly cover QA pairs asked on community sites
ANTIQUE	Comprises 2626 non-factoid questions collected from Yahoo! Answers community question answering websites and their 34011 corresponding annotated answers. Out of these non-factoid questions, approximately 900 are of why-type which are not enough for research in only why-QAS

4- Open-Domain Why-QA Dataset Preparation

This section discusses the process of collecting why-QA pairs and processing them to create a final dataset.

4-1-Customized Crawling and Scraping

Five web sites viz. Answers.com [15], Yahoo! Answers [13], Suzan Verberne's open-source dataset [16], Quora [17] and Ask.com[18] are visited by web crawler and Scrapy [19] which is an open source framework written in Python is utilized to scrap or extract required data from these websites. A customized spider is implemented which extracts all questions beginning with 'Why'. BeautifulSoup [20] is used to parse the HTML source of web page and questions are extracted from the 'href' tag of HTML source. The process resulted in a collection of Why-questions having different forms like 'Why is/was/were', 'Why do/does/did', 'Why should/would' and 'Why NP/PN' with their negative variants.

The functioning of web crawling and scraping is depicted in the form of Flowchart illustrated in Figure 1 and discussed further.

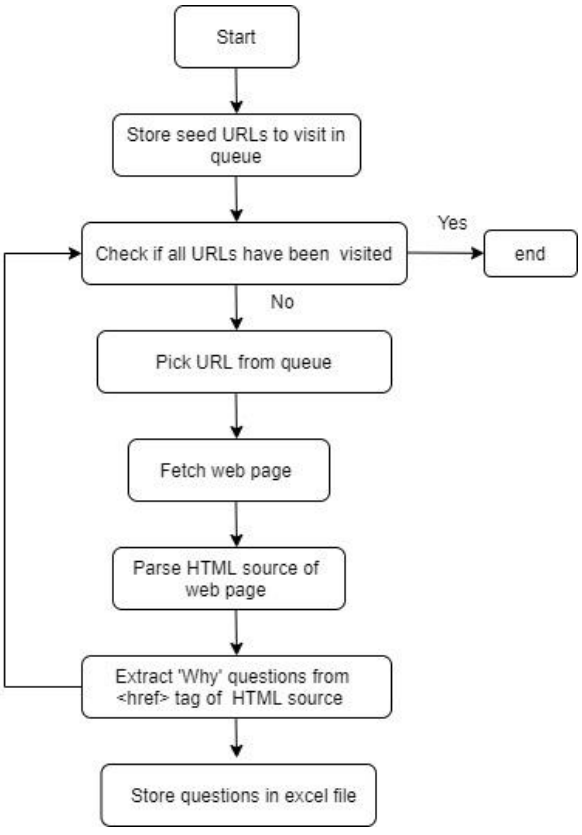


Fig. 1. Flowchart describing functioning of web crawler

Web crawler also known as spider or search engine bot is a part of search engines for example, Google, Bing, Yahoo etc. It is a software program that accesses a website, downloads and indexes information content available on the web page. To make automated browsing, the crawler is designed to repeat accessing and downloading content depending on the programming instructions included in crawler.

Since content on Internet is expanding and updating daily, it is important to index the webpages so that it can be easily accessible and downloadable based on users' query. Web crawler starts from a seed URL and crawl the webpages, follow hyperlinks to those URLs, thus periodically visit pages in order to index the updated content on webpages. Sometimes the webpages which are required to crawl depends on number of hyperlinks to the page, amount of visitors on page and other factors determining the importance of page corresponding to the required content.

Web crawling and scraping are two distinguished terms. Web crawlers continuously visit web links and download contents on web pages. Web crawlers obey robots.txt (robots exclusion protocol) file provided by web server of particular web page. This file specifies the rules for which pages and links to pages can be crawled. On the other hand, Web scraping is more targeted than web crawling and visit only specific webpages to download their content without any permission from the web server.

Figure 1 describes the process of crawling where crawler starts with seed URLs which are question answering sites. If all sites are visited, the process is stopped otherwise crawler visits each webpage and retrieves its HTML source. Here the crawler is designed to download questions beginning with 'why', thus it is extracted by scraping <href> tag of HTML source of web page. The process is repeated for all QA websites available and the questions are stored in excel file.

4-2-Processing of Why-Questions

A collection of 2000 why-questions are downloaded and stored in excel file. The questions are further processed to rectify grammatical and spelling mistakes. The why-questions of the form starting with 'Why' and ending with '?' are retained in the dataset. Question comprising more than one questions are separated into individual questions, coreference resolution [21] is applied to filter out significant meaningful why-type questions. Coreference resolution is NLP (Natural Language Processing) task which finds linguistic expressions such as pronouns in the question and replace them with real-world entity such as noun phrases which helps to understand the appropriate meaning of the question and help determine the main focus of question.

4-3- Different Fields in the Dataset

With each question, different attributes are also maintained in the dataset which are:

a) *Data Source of question:* There are five web sites from which web spider has extracted why-type questions. The data source from which the why-question is collected is stored as an attribute 'source_ques' in dataset.

b) *Category of the question:* With each why-type question, its category is stored in dataset. The questions collected from Answers.com and Yahoo have their category assigned like Science, Math, History, Technology and so on whereas questions collected from other web sites are assigned category by applying LDA process [22,23] which is one of the topic modelling approach. Latent Dirichlet Allocations abbreviated as LDA models Dirichlet distributions to classify questions to a particular topic. This category is stored with field name 'categ_ques' in the dataset.

c) *Relevant answer candidates:* For each why-type question, five answer candidates are maintained in the dataset. Only if the question is answered by an expert on the considered five question answering websites, its corresponding answer is saved as an answer candidate otherwise the answer candidates are retrieved by posing the question on Google search engine. The question is queried on the Google which extracts relevant web pages from which first five are studied manually to retrieve appropriate answer passages treated as answer candidate.

Since the answer to why-question involves explanation to the causation asked in the question, appropriate answer candidate is retrieved by determining the answer boundary around causal cue phrases [24] such as 'because', 'since', 'due to', 'in order to', 'therefore', 'as a result' etc. The causal phrases determine explicit causality involved in text. The why-type questions reflect the effect part and its cause part is reflected from the answers to why-type questions.

d) *Document title and link:* Each why-question is accompanied with the document title and its link which is the title and the link of web page from which an answer candidate is retrieved. It is saved as an attribute 'doc_title_link' in the dataset.

The snapshots of the generated dataset are illustrated in Figure 2 and Figure 3 below:

Data source of question	category of question	Relevant Answer Candidates	Document Title	Document Link
Quora	Travel	The Taj Mahal is located on the right bank of the Yamuna River in a vast Mughal garden that encompasses nearly 17 hectares, in the Agra District in Uttar Pradesh. It was built by Mughal Emperor Shah Jahan in memory of his wife Mumtaz Mahal with construction starting in 1632 AD and completed in 1648 AD, with the mosque, the guest house and the main gateway on the south, the outer courtyard and its cloisters were added subsequently and completed in 1653 AD. The existence of several historical and Quaranic inscriptions in Arabic script have facilitated setting the chronology of Taj Mahal.	Taj Mahal	https://whc.unesco.org/en/list/252/
Quora	Travel	"The Taj Mahal, meaning "Crown of the Palace", is an ivory-white marble mausoleum on the south bank of the Yamuna river in the Indian city of Agra. It was commissioned in 1632 by the Mughal emperor, Shah Jahan (reigned from 1628 to 1658), to house the tomb of his favourite wife, Mumtaz Mahal. The tomb is the centerpiece of a 17-hectare (42-acre) complex, which includes a mosque and a guest house, and is set in formal gardens bounded on three sides by a crenellated wall."	Why was the Taj Mahal built?	https://www.quora.com/What-is-the-real-reason-The-Taj-Ma
Quora	Travel	Often described as one of the wonders of the world, the stunning 17th Century white marble Taj Mahal was built by Mughal emperor Shah Jahan as a mausoleum for his beloved wife Mumtaz Mahal, who died in childbirth.	Taj Mahal: Was India's 'monument to love' built out of guilt?	https://www.bbc.com/news/world-asia-india-27970691
Quora	Travel	The Taj Mahal was built by the Mughal emperor Shah Jahan (reigned 1628–58) to immortalize his wife Mumtaz Mahal ("Chosen One of the Palace"). She died in childbirth in 1631, after having been the emperor's inseparable companion since their marriage in 1612	Taj Mahal	https://www.britannica.com/art/Shah-Jahan-period-architect
Quora	Travel	The Taj Mahal is an ivory-white marble mausoleum on the southern bank of the river Yamuna in the Indian city of Agra. It was commissioned in 1632 by the Mughal emperor Shah Jahan (reigned from 1628 to 1658) to house the tomb of his favourite wife, Mumtaz Mahal; it also houses the tomb of Shah Jahan himself.	Taj Mahal	https://en.wikipedia.org/wiki/Taj_Mahal

Fig. 2. Snapshot 1 of dataset

Why is Venus hottest planet of solar system?	Quora	Science & Mathematics	Venus is the second planet from the Sun and our closest planetary neighbor. Similar in structure and size to Earth, Venus spins slowly in the opposite direction from most planets. Its thick atmosphere traps heat in a runaway greenhouse effect, making it the hottest planet in our solar system with surface temperatures hot enough to melt lead.	Venus	https://mobile.arc.nasa.gov/public/explore/missions/pages/galileo/extent/venus.html
Why is Venus hottest planet of solar system?	Quora	Science & Mathematics	Venus is the hottest planet in the solar system. Although Venus is not the planet closest to the sun, its dense atmosphere traps heat in a runaway version of the greenhouse effect that warms Earth. As a result, temperatures on Venus reach 880 degrees Fahrenheit (471 degrees Celsius), which is more than hot enough to melt lead. Spacecraft have survived only a few hours after landing on the planet before being destroyed.	Venus: The hot, hellish & volcanic planet	https://www.space.com/64-venus-second-planet-from-the-sun-brightest-planet-in-solar-system.html
Why is Venus hottest planet of solar system?	Quora	Science & Mathematics	Even though Mercury is the closest planet to the Sun, Venus is the hottest planet in our solar system. This is because Mercury has almost no atmosphere, while Venus has a very thick atmosphere. This causes all the heat to be radiated back into space on Mercury. However, the heat is trapped on Venus, with the average temperature being 462°C.	Why is Venus the hottest planet in our solar system?	https://indiaarts.com/knowledge/why-is-venus-the-hottest-planet-in-our-solar-system-1688260417879
Why is Venus hottest planet of solar system?	Quora	Science & Mathematics	Venus is so hot because it is surrounded by a very thick atmosphere which is about 100 times more massive than our atmosphere here on Earth. As sunlight passes through the atmosphere, it heats up the surface of Venus. Most of this heat cannot escape back into space because it is blocked by the very thick atmosphere of Venus. The heat becomes trapped and builds up to extremely high temperatures. This trapping of heat by the atmosphere is called the greenhouse effect because it is similar to how the glass in a greenhouse traps heat. The greenhouse effect on Venus causes the temperatures at its surface to reach 864 degrees Fahrenheit (462 degrees Celsius), making Venus the hottest planet in the entire Solar System!	Why is Venus so hot? Surface Of Venus: Why Is Venus The Hottest Planet?	https://coolcosmos.ipac.caltech.edu/ask/38-Why-is-Venus-so-hot-
Why is Venus hottest planet of solar system?	Quora	Science & Mathematics	Venus is the hottest planet in our solar system because it is covered by a thick layer of clouds composed of carbon dioxide and other gases, which prevent the heat from the sun from escaping back into outer space. This is why the planet continues absorbing the heat from the sun and becomes increasingly hot.		https://www.sciencedbc.com/nature/universe/surface-of-venus-why-is-venus-the-hottest-planet.html

Fig. 3. Snapshot 2 of dataset

5- Results and Discussion

The crawler designed which visits each question answering web sites, extracts HTML source and retrieve why-type questions available on the QA sites, resulted in a collection of around 2000 why-type questions. The dataset is further enhanced by incorporating data source, category, five answer candidates, document link of the respective answer candidate. The dataset is analyzed on the distribution of topics and distribution of different forms of why-questions. The analysis is performed in order to contrast the dataset with available open-source datasets and further help the researchers get an idea of the variations involved in the dataset collection.

5-1- Variation of Topics in Why-Questions of Dataset

The why-questions in dataset have variation in their topics. Various different categories are visualized like 'Education', 'Politics', 'Health', 'Science & Mathematics', 'Family & Relationships', 'Travel', 'Sports', 'Electronic Products', 'Food, Drink & Dining Out' and 'Entertainment & Music' etc. The distribution of questions belonging to these major categories and others is illustrated in Figure 4 below:

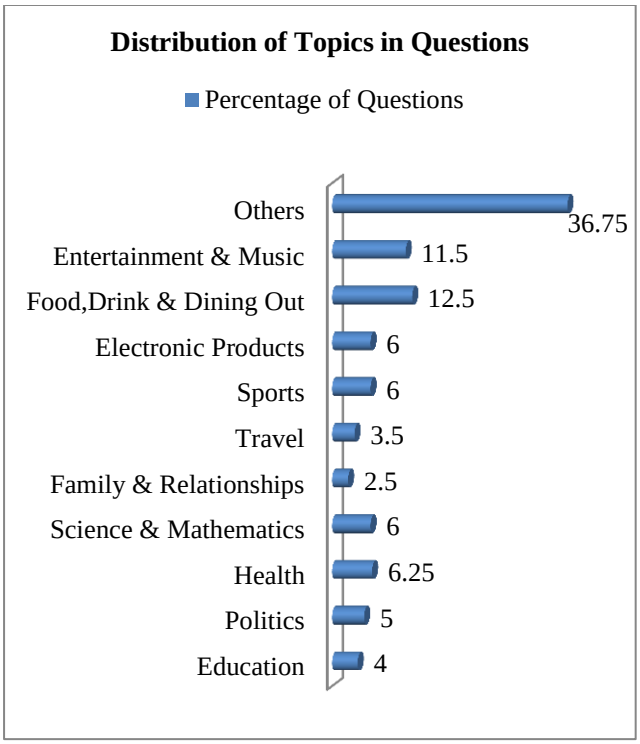


Fig. 4. Distribution of topics in why-questions

5-2- Distribution of Forms of Why-Questions

As mentioned in subsection 3.1, there are different forms of why-questions viz. 'Why is/was/were', 'Why do/does/did', 'Why should/would' and 'Why NP/PN' with their negative variants. Figure 5 illustrates the distribution of subcategories of why-type questions.

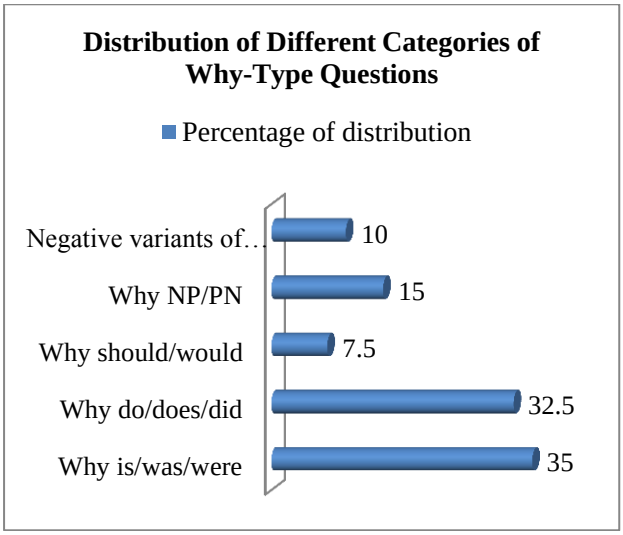


Fig. 5. Distribution of different forms of why-questions

6- Usage of Why-QA Dataset

This section highlights expected usage of why-QA dataset, some of which we have utilized in our research.

a) *Question Classification:* An open-domain why-QA dataset is utilized for classifying and assigning question and answer subtype. A taxonomy of why-type questions and their answers are proposed in the research [25,26, 36] which extracts lexical features of questions and assigns a question and answer type to them.
b) *Question to Query Reformulation:* The reformulated question helps in better document retrieval and thus answer candidate extraction. Rules are formulated for each type of why-question to convert them into appropriate user-oriented query which when posed on search engine helps in retrieving the accurate and relevant documents [37].
c) *Answer Re-Ranker:* With each why-type question, there are five possible answer candidates stored in the dataset. These answer candidates are re-ranked on the basis of similarity values and their relatedness with question [38].

4- Conclusions and Future Work

The paper proposes a dataset containing why-type questions and their answer candidates. Since the questions are sampled from various question answering websites such as Yahoo! Answers, Quora etc., the questions address the real-world problems usually faced by the user. The questions in dataset are collected with the help of web crawler in its original form, thus easier to predict properties and nature of why-type questions.
We believe that the developed open-domain why-QA dataset unfolds new avenues for research in improving performance of non-factoid QAS. With this, it also motivates to foster the techniques required for other applications like recommendation systems, chatbots, virtual assistants and many more.

References

[1] S. Verberne, L.W.J. Boves, N.H.J. Oostdijk and P.A.J.M. Coppen, "Data for question answering: the case of why", 2006.
[2] shuzi, "GitHub - shuzi/insuranceQA: A question answering corpus in insurance domain", 2015. [Online]. Available: https://github.com/shuzi/insuranceQA. [Accessed Feb. 9, 2021].
[3] D. Cohen, L. Yang,, and W. B. Croft, "Wikipassageqa: A benchmark collection for research on non-factoid answer passage retrieval", In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 1165-1168.
[4] A. Dulceanu, T. Le Dinh, W. Chang, T. Bui, D.S. Kim, M.C.Vu, and S. Kim, "PhotoshopQuiA: A corpus of non-factoid questions and answers for why-question answering", In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018, pp. 2763-2770.
[5] Stack Overflow - Where Developers Learn, Share, & Build Careers. [Online]. Available: https://stackoverflow.com/. [Accessed Feb. 9, 2021].
[6] Adobe Support Community. [Online]. Available: https://forums.adobe.com/welcome. [Accessed Feb. 9, 2021].
[7] Graphic Design Stack Exchange. [Online]. Available https://graphicdesign.stackexchange.com. [Accessed Feb. 9, 2021].
[8] Super User Stack Exchange [Online]. Available https://superuser.com. [Accessed Feb. 9, 2021].
[9] Adobe Photoshop Family [Online]. Available https://feedback.photoshop.com. [Accessed Feb. 9, 2021].
[10] K. Jiang,, D. Wu and H. Jiang., "FreebaseQA: a new factoid QA data set matching Trivia-style question-answer pairs with freebase", In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), June 2019, pp. 318-323.
[11] T. Kočiský, J. Schwarz, P. Blunsom, C. Dyer, K. M. Hermann, G. Melis and E. Grefenstette, "The narrativeqa reading comprehension challenge", Transactions of the Association for Computational Linguistics, vol. 6, pp. 317-328, 2018.
[12] H. Hashemi, M. Aliannejadi, H. Zamani and W.B. Croft, "ANTIQUE: A non-factoid question answering benchmark.". In European Conference on Information Retrieval, Springer, Cham, 2020, pp. 166-173.
[13]Yahoo! answers [Online]. Available https://answers.yahoo.com/. [Accessed Feb. 9, 2021].
[14] A. Colas, S. Kim, F. Dernoncourt,, S. Gupte, D.Z. Wang and D.S. Kim, "TutorialVQA: Question Answering Dataset for Tutorial Videos" [Online]. Available arXiv preprint arXiv:1912.01046, 2019.
[15] Answers [Online]. Available https://www.answers.com/. . [Accessed Feb. 9, 2021].
[16] S. Verberne, "Data Download," [Online]. Available: http://sverberne.ruhosting.nl/wordpress/research/data-download/. [Accessed Feb. 9, 2021].
[17] Quora [Online]. Available https://www.quora.com/ . [Accessed Feb. 9, 2021].
[18] Ask [Online]. Available https://www.ask.com/. [Accessed Feb. 9, 2021].

[19] Scrapy [Online]. Available https://scrapy.org/. [Accessed Feb. 9, 2021].
[20] BeautifulSoup [Online]. Available https://pypi.org/project/beautifulsoup4/. [Accessed Feb. 9, 2021].
[21] A. Rahman and V. Ng, "Coreference resolution with world knowledge", In Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, 2011, pp. 814-824.
[22] A. Anandkumar, D. P. Foster, D. Hsu, S.M. Kakade and Y.K. Liu, "A spectral algorithm for latent dirichlet allocation", Algorithmica, vol. 72, no. 1, 2015, pp. 193-214.
[23] D. M. Blei, A.Y. Ng and M. I. Jordan, "Latent dirichlet allocation.", Journal of machine Learning research, 2003, pp. 993-1022.
[24] D.S. Chang and K.S. Choi, "Causal relation extraction using cue phrase and lexical pair probabilities", In International Conference on Natural Language Processing, Springer, Berlin, Heidelberg, 2004, pp. 61-70.
[25] M. Breja and S.K. Jain, "Why-type Question Classification in Question Answering System", In FIRE (Working Notes), 2017, pp. 149-153.
[26] M. Breja and S.K. Jain, "Analysis of Why-Type Questions for the Question Answering System", In European Conference on Advances in Databases and Information Systems, Springer, Cham, 2018, pp. 265-273.
[27] H. Fu and Y. Fan, "Music information seeking via social Q&A: An analysis of questions in music StackExchange community", In Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, 2016, pp. 139-142.
[28] Financial Opinion Mining and Question Answering [Online] Available https://sites.google.com/view/fiqa/home/. [Accessed Feb. 9, 2021].
[29] A.F.U.R. Khilji, R. Manna, S.R. Laskar, P. Pakray, D. Das, S. Bandyopadhyay and A. Gelbukh, "Question classification and answer extraction for developing a cooking QA system", Computación y Sistemas, vol. 24, no. 2, 2020, pp. 921-927.
[30] V. Koeman, L. A. Dennis, M. Webster, M. Fisher and K. Hindriks, "The Why did you do that?" Button: Answering Why-questions for end users of Robotic Systems", In 7th International Workshop on Engineering Multi-Agent Systems (EMAs 2019), 2019, pp. 152-172.
[31] Yahoo! Language Data [Online]. Available https://webscope.sandbox.yahoo.com/catalog.php?datatype=l. [Accessed Feb. 9, 2021].
[32] E. Hovy, U. Hermjakob, and D. Ravichandran, "A question/answer typology with surface text patterns", In Proceedings of the Hum an Language Technology conference (HLT), San Diego, CA, 2002.
[33] A. Mishra and S.K. Jain, "A survey on question answering systems with classification", Journal of King Saud University-Computer and Information Sciences, vol. 28, no. 3, 2016, pp.345-361.
[34] The Baron [Online]. Available https://www.thebaron.info/archives/technology/reuters-technical-development-chronology-1991-1994, [Accessed March, 02, 2021]
[35] G. Smith, "Newspapers on CD-ROM", In Serials The Journal for the Serials Community, vol. 5, no. 3, 1992, pp. 17-22.
[36] M. Breja and S.K. Jain, "Analyzing Linguistic Features for Classifying Why-Type Non-Factoid Questions", International Journal of Information Technology and Web Engineering (IJITWE), vol. 16, no. 3, 2021, pp.21-38.
[37] M. Breja and S.K. Jain, "Why-Type Question to Query Reformulation for Efficient Document Retrieval", International Journal of Information Retrieval Research (IJIRR), vol. 12, no. 1, 2022, pp.1-18.
[38] M. Breja and S.K. Jain, "Analyzing Linguistic Features for Answer Re-Ranking of Why-Questions", Journal of Cases on Information Technology (JCIT), vol. 24, no. 3, 2022, pp.1-16.
[39] M. Breja and S.K. Jain, "A survey on non-factoid question answering systems", International Journal of Computers and Applications,2021, pp.1-8.
[40] Y. Niu , "Analysis of semantic classes: toward non-factoid question answering". University of Toronto, 2007 .

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 48-56
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Mathematical Modeling of Flow Control Mechanism in Wireless Network-on-Chip

Farhad Rad[1*], Marzieh Gerami[1]

[1]. Department of Computer Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran
[2]. Department of Computer Engineering, ShahreKord Branch, Islamic Azad University, ShahreKord, Iran

Received: 21 Aug 2021/ Revised: 04 Jan 2022/ Accepted: 12 Feb 2022

Abstract

Network-on-chip (NoC) is an effective interconnection solution of multicore chips. In recent years, wireless interfaces (WIs) are used in NoCs to reduce the delay and power consumption between long-distance cores. This new communication structure is called wireless network-on-chip (WiNoC). Compared to the wired links, demand to use the shared wireless links leads to congestion in WiNoCs. This problem increases the average packet latency as well as the network latency. However, using an efficient control mechanism will have a great impact on the efficiency and performance of the WiNoCs. In this paper, a mathematical modeling-based flow control mechanism in WiNoCs has been investigated. At first, the flow control problem has been modeled as a utility-based optimization problem with the wireless bandwidth capacity constraints and flow rate of processing cores. Next, the initial problem has been transformed into a dual problem without limitations and the best solution of the dual problem is obtained by the gradient projection method. Finally, an iterative algorithm is proposed in a WiNoC to control the flow rate of each core. The simulation results of synthetic traffic patterns show that the proposed algorithm can control and regulate the flow rate of each core with an acceptable convergence. Hence, the network throughput will be significantly improved.

Keywords: Wireless Network-on-Chip; Flow Control Mechanism; Optimization; Gradient Projection Method; Utility Function.

1- Introduction

One of the simplest communication structures of the components inside the chip is the bus. The main drawback of bus-based systems is the sharp decline of the system performance with the increasing number of processing cores. Increasing the number of cores integrated into a chip will transform the communication bus between the cores into a bottleneck for the system. Also, in this structure, having a mechanism as an arbitrator for control and fair access to the joint bus seems necessary. The inflexibility of this communication structure has led to a significant challenge in systems-on-chip.

The network on the chip is a communication structure between the cores within the chip that has overcome the defects and challenges of the few structures. On-chip networks, while being scalable, should provide low power consumption and adequate bandwidth for communication between tens or hundreds of processing cores within a chip. A major challenge of the network on the chip, despite the reduction in transistor dimensions in recent years, is the issue of delayed wired connections within the chip, which has grown exponentially (ITRS report) [1].

In recent years, new communication structures include 3DNoCs, photonic NoCs, and network on chip with radio (wireless) connections have been proposed [2]. Using these emerging technologies in the advanced integrated circuit industry has been able to reduce the delay and power consumption of the chip.

The design and manufacture of on-chip integrated silicon antennas were presented in the last decade. These antennas have been used with a frequency equivalent to ten to one hundred GHz inside a chip [3]. However, increasing the operating frequency of processing cores and the high bandwidth of wireless links for inter-core communication will turn these links into hotspots in the network (network congestion will increase delay and reduce network efficiency). Therefore, the existence of a mechanism to control congestion and flow control to reduce network traffic and increase chip performance is essential.

✉ **Farhad Rad**
Fa.rad@iau.ac.ir

According to the studies conducted in the field of wireless networks-on-chip, the problem of modeling the flow rate control in the wireless network on the chip has not been done so far. For this purpose, in this paper, first, the flow rate control modeling in the wireless network on the chip will be examined. To formulate the problem, we have used the maximization of a utility function. Then, we solved the problem with Lagrange multipliers and gradient projection and presented an algorithm to control the flow rate of processing cores. The proposed algorithm has been able to regulate and control the flow rate of processing cores with appropriate convergence speed.
This paper consists of the following parts:
Section 2 has presented the details of the wireless network-on-chip. The network model as well as the mathematical modeling of the flow control problem along with the proposed algorithm have been presented in Section 3. The results of the implementation have been given in Section 4. Finally, the conclusion has been described in Section 5.

2- A review of Previous Works

The use of wireless in-chip connections has introduced a new structure called the wireless network on chip (WiNoC) [4]. In addition to high flexibility, this network has increased communication bandwidth and eliminated the problems of noise and power consumption of metal wires. This new structure on the chip has been generally studied in three layers. Physical layer, data link layer, and network layer. The following is a brief introduction to each section.

2-1- Physical Layer in the WiNoC:

In a WiNoC, antennas are designed according to the operating frequency of the circuits and then used in conjunction with other wireless transmission devices such as modulator circuits and telecommunication filters on the chip. Since the bit rate error in wireless on-chip transmission will have a significant effect on the reliability of the system, hence the medium of wave propagation must be completely examined. The use of different frequency boards has divided the WiNoC into four general categories. Figure 1 shows these four frequency ranges that can be used in the design of WiNoC.

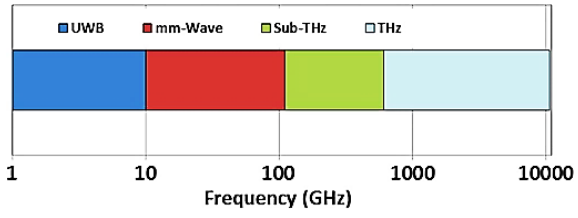


Fig. 1. Division of the frequency of the WiNoC [4]

In the following, we will briefly review the work done in the field of designing antennas compatible with each of the frequency ranges:

2-1-1- Ultra-Wide Band (UWB):

Reference [5] has proposed the architecture of WiNoC based on the frequency of the ultra-wideband. In this architecture, a radio pulse transmitter-receiver with a carrier signal has been used for wireless communication. The transmitter uses an integrated, CMOS-compliant Gaussian monocycle pulse to generate low-power, low-driving pulses. The antenna used in this architecture is a 2.9mm long Meander Dipole Antenna (MDA) that can transmit data up to a radius of 1 mm. The transmission rate of this antenna is 1.16 Gbps for a channel with a frequency of 3.6 GHz.

2-1-2- Mm-Wave Band:

reference [3] has proposed the design of a metal zigzag antenna. This antenna will have little effect on the rotation (relative angle between the transmitting antenna and the receiving antenna) and the received signal strength. The millimeter frequency band was also able to provide bandwidth between ten and one hundred GHz without signal attenuation.

2-1-3- Sub-THz:

The design of small antennas with simple transceivers that could operate at frequencies between 100 and 1000 GHz has been proposed for the first time in reference [6]. These antennas can cover a communication radius of 10 to 20 mm on the chip. In this architecture, the possibility of simultaneous use of 16 non-overlapping channels with frequencies of 100 to 500 GHz for the structure of the WiNoC has been proposed.

2-1-4- THz:

by increasing the signal frequency to the THz range, one can expect much smaller antennas to be designed that can occupy less surface area on the chip. Reference [7] has proposed the design and fabrication of antennas using carbon nanotubes. The properties of carbon nanotubes make them suitable for making antennas on high-frequency chips.
In addition to the above, the design of graphene-based plasmonic antennas using electromagnetic waves up to the range of ten terahertz for wireless in-chip communication has also been introduced and has recently been considered [8]. In this paper, the authors claim that the many benefits of using graphene will bring great success in wireless in-chip communication. In the future, the design of these antennas will provide the possibility of all-broadcast and multi-broadcast on-chip communication. Figure 2 shows a view of the network-on-chip with graphene antennas.

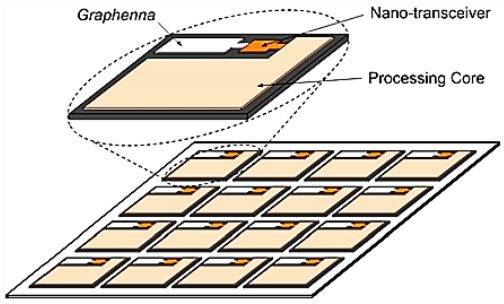


Fig. 2. Graphene antennas in the structure of WiNoC [8]

2-2- Data Link Layer in the WiNoC:

To increase the efficiency and maximum use of the bandwidth of communication channels in the WiNoC, having a media access control mechanism (MAC) is also needed. Designing a MAC protocol on a wireless chip should be able to impose the minimum of power and consumption overhead on the system. In reference [5] a shared channel access protocol was proposed. In this design, the network is divided into several areas with the least signal interference. Hence, to access shared wireless channels a central arbitration policy is used.

In [9], a protocol has been implemented based on token flow control. In this design, any router equipped with a wireless antenna that has the token can send or receive packets on the wireless channel. After sending the packets, this token will flow in turning form for the use of other routers. By increasing the number of wireless cores in the NoC, this mechanism leads to an increase in token flow time between cores and a decrease in network performance.

Reference [10] has proposed an access control mechanism called RACM for WiNoC. In this design, each radio hub (wireless router) has been connected to a processing core on a ring-shaped structure. This mechanism dynamically splits the token capture time in a router that has no data to send between all stations that have used their maximum time to send data. The simulation results show a 30% reduction in communication delay and a 25% improvement in stored energy in this design.

2-3- Network Layer in the WiNoC:

For maximum performance in the WiNoC, it is necessary to have deadlock-free and fault-tolerant routing algorithms. A fully adaptive fault-tolerant routing algorithm (FAFTR) has been proposed and investigated in reference [11]. In addition to finding the shortest path, the proposed distributed algorithm can be free of deadlocks. In this algorithm, the deadlock has been prevented by applying rotation restrictions.

A deadlock-free routing algorithm based on the static routing table and the XY algorithm, with a sub-THz frequency band in the mesh structure, has been also presented in the reference [12]. Reference [13] has presented an algorithm that uses the buffer state of wireless routers to decide on the shortest path. This algorithm has been able to prevent traffic in wireless routers.

The use of routing algorithms along with Media Access Control (MAC) mechanisms as well as core flow rate control solutions can provide better network performance in the WiNoC while properly distributing traffic throughout the network.

Transmitter flow control algorithms can also manage the competitive conditions created between cores for shared resources (such as channel bandwidth or buffers) and provide fair resource allocation. Hence, two types of flow control in the NoC can be considered:

2-3-1- Switch-to-Switch (Router) Mechanisms:

in this design, control signals between adjacent routers have been used locally to regulate the flow rate of traffic. Examples include credit-based flow control, on-off flow control, acknowledgment/non-acknowledgment protocol, and handshaking signal-based flow control. These methods do not require direct communication between the transmitter and receiver to exchange control signals and do not impose much control overhead on the system. However, as the number of cores increases and the size of the network increases, before the path congestion information in the form of control signals can reach the origin at the appropriate time, the packets sent by the transmitter will inevitably increase the network congestion [14].

2-3-2- End-to-End Flow Control Mechanisms:

in this mechanism, the transmitter cores will be responsible for the rate of production and injection of their flows into the network. This method eliminates packet overload and network congestion. In the network on chip, unlike computer networks, it is not possible to delete and lose packets, therefore, there is no need for confirmation signals to send and receive packets safely. Therefore, end-to-end flow control methods will not impose much control overhead on the system and each core will be independently responsible for controlling and regulating the transmitted flow rate [15].

In the following, we will first review the works done on the problem of congestion control/flow control in the NoC. Then, the works done on the flow rate control in the WiNoC have been presented. Reference [14] has proposed a prediction-based congestion control algorithm. In this design, each router decides on the congestion that may occur in the future based on the size and status of its gateway buffer. This design regulates the number of packets in the network by controlling the flow rate of injection of packets into the network.

In reference [16], link utilization has been used as a measure of congestion. A controller determines the appropriate workload for the current transmitters at best.

In the reference [17] of the CAP-W algorithm, a combined routing algorithm has been used to solve the problem of local

congestion in the hierarchical wired and wireless structure. The proposed routing algorithm has tried to control congestion in the area by distributing traffic throughout the network by changing the status of a deterministic routing to adaptive routing and also proposing mapping tasks to freer cores.

Distributed congestion control method has been proposed in the reference [18]. To detect congestion, the number of packets in the queue buffers has been used. The main disadvantage of this method is that due to the limited space of the buffer of each router, it is possible to detect the fault, especially when the network congestion is relatively high.

A congestion control mechanism based on the flow rate utility function and the total delay was presented in references [19-21].

In a WiNoC, two types of competitive flow have been proposed [22-23]. The authors have presented a distributed flow control mechanism with a buffer management strategy to improve network performance on the wireless chip. In [23] the buffer information of each wireless router is used to prevent congestion. On the other hand, using the deterministic routing algorithms lead to Head-of-Line (HoL) blocking when load traffic is high [30-32]. Hence, some studies used congestion control mechanisms to eliminate HoL and reduce congestion in the wireless routers congestion-aware routing algorithms [33-34].

3- Mathematical Modeling of Flow Rate Control in WiNoC

In this section, we will first describe the network model used and then the flow rate control problem has been modeled.

3-1- Model of WiNoC

The NoC is based on a two-dimensional mesh connection with several routers equipped with a wireless antenna in the center of each 3 × 3 sections. Figure 3 shows a view of this structure.

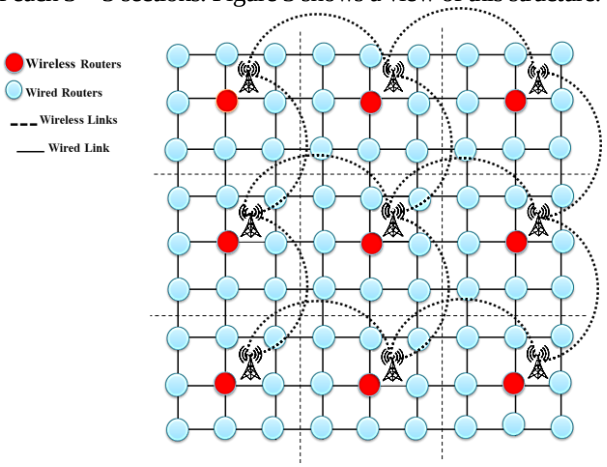


Fig. 3. Graphic structure of the proposed network

The wireless structure is based on a mechanism of Frequency Division Multiple Access (FDMA) and wireless single-hop transmission. Each wireless router can access the other routers in the adjacent subdivision using a single hop. In this structure, the bandwidth of wireless channels has been assumed to be fixed and without bit error. All wireless routers (red nodes) are in direct view of each other and each pair of them operates in a separate frequency band (channel) from the other pairs. The switching method is wormhole switching. The X-Y routing algorithm is used to route packets (first the packet moves in the X direction and then in the Y direction) which is a deadlock-free algorithm. Routing between each transmitter and receiver pair located in 3 × 3 areas is performed using the X-Y routing algorithm. If the transmitter and receiver are not in a 3 × 3 section, then a combination of the X-Y algorithm and wireless links are used to send packets between the source and destination. A matrix A has been established by using this routing pattern. This matrix will show the routing pattern between the transmitter and receiver. In the implemented structure, the size of matrix A is 64 × 36 (64 rows and 36 columns for 6 × 6 networks). Since one of the features of routing algorithms of the NoC is to prevent deadlocks, therefore, to prevent this phenomenon, the octagon turn model [24] has been used in the proposed method. This model uses four clockwise rotations and four counterclockwise rotations to prevent deadlock and create cycles. The rules of this model have been stated as follows:

• The flit of a package will not be allowed to turn in the following four clockwise directions:

W→SE, N→SW, E→NW, and S→NE

• The flits of a package will not be allowed to turn in the following four counterclockwise directions:

NE→S, NW→E, SW→N, and SE→W

Figure 4 shows the eight ports (directions in which the flits of a packet can be sent or received) from a router equipped with a wireless antenna. Table 1 also shows the symbols used in the proposed modeling.

Table 1. Symbols used in the proposed model

Definition	Symbol
Core set	$k \in K = \{1,2,\ldots\ldots,K\}$
A set of wired and wireless links	$l \in L = \{1,2,\ldots,L\}$
Injection rate per core	$x = (x_k, k \in K)$
The bandwidth of each link	$c = (c_l, l \in L)$
A set of links through which the core flow passes	$L(k) \subseteq L$
A set of cores that pass through link l	$K(l) \subseteq K$
Traffic pattern matrix	$A = (A_{lk})_{L*K}$

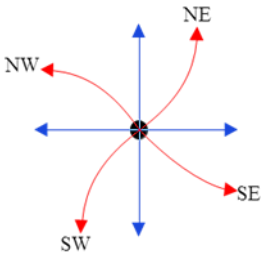


Fig. 4. Wireless router with eight transmission and receiving directions

3-2- Flow Control Modeling Problem

In this section, the flow control problem in the WiNoC has been formulated using mathematical modeling. To control and regulate flow rates, the problem has been written in a utility-based function based on the limited bandwidth capacity of links as well as flow rates. This method aims to maximize the flow rate of all processing cores taking into account the mentioned limitations. The concept of utility function was inspired by economics science and shows the degree of satisfaction a consumer or seller of a product at his disposal. In the proposed model, a utility function called U is defined for each transmitter (k) based on the flow rate (x). The purpose of modeling is to maximize the sum of all rates of in-chip cores to achieve maximum optimum chip performance. The function has been considered as an incremental, derivative, and concave function [25]. Figure 5 shows the steps of modeling operations.

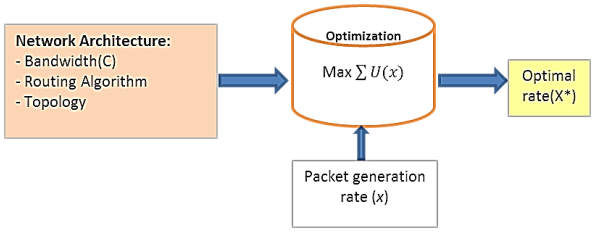


Fig. 5. Steps of flow control problem modeling

Equation 1 has shown that the sum of the flow rates through a link cannot be greater than the bandwidth capacity of the link. The symbol A is the link traffic pattern matrix written by the routing algorithm implemented in the target network.

$$A * x \leq c \tag{1}$$

Now, to control the flow rate of the cores to maximize total flow rates, the following relationships can be written:

$$max \sum_{k \in K} U_k(x_k) \tag{2}$$

subject to:

$$A \,.\, x \leq c \tag{3}$$

$$x_k > 0 \qquad \forall k \in K \tag{4}$$

To solve the written nonlinear optimization problem, Lagrange multipliers can be used to rewrite the problem:

$$L(x, \rho^\alpha) = \sum U_k(x_k) - \rho^\alpha (A.x - c) \tag{5}$$

$$L(x, \rho^\alpha) = \sum_{k \in K} U_k(x_k) - \sum_{l \in L} \rho_l^\alpha \left(\sum_{k \in K} A_{lk} x_k - c_l \right) \tag{6}$$

$$L(x, \rho^\alpha) = \sum_{k \in K} U_k(x_k) - \left(\sum_{l \in L} \rho_l^\alpha \sum_{k \in K} A_{lk} x_k \right) + \sum_{l \in L} \rho_l^\alpha c_{l=}$$

$$\sum_{k \in K} U_k(x_k) - \sum_{k \in K} x_k \sum_{l \in L} \rho_l^\alpha A_{lk} + \sum_{l \in L} \rho_l^\alpha c_l \tag{7}$$

$$\lambda_l^\alpha = \sum_{l \in L} \rho_l^\alpha A_{lk} = \sum_{l \in L(k)} \rho_l^\alpha \tag{8}$$

λ_l^α : The total cost of links through which the same flow passes. (The cost that a core must pay to pass its flow through these links [26]).

According to the dual theory, for any problem, dual convex optimization can also be written. The dual form of Equation 2 has been written in the following form [27]:

$$L(x, \rho^\alpha) = \sum_{k \in K} U_k(x_k) - (\lambda^\alpha x) + \rho^\alpha c \tag{9}$$

$$D: \min_{\rho^\alpha \geq 0} D(\rho^\alpha)$$

where

$$D(\rho^\alpha) = \max_x L(x, \rho^\alpha)$$

$$= \max_x \sum_{k \in K} (U_k(x_k) - \lambda_k^\alpha x_k) + \sum_{l \in L} \rho_l^\alpha c_l \tag{10}$$

Considering the concavity and derivability of the utility function, the final and global value of the flow rate can be obtained from the following equation:

$$D(\rho^\alpha) = \sum_{k \in K} \max_{x_k \in [m_x, M_x]} \{U_k(x_k) - \lambda_k^\alpha x_k\} + \sum_{l \in L} \rho_l^\alpha c_l \tag{11}$$

$$U_k(x_k) - \lambda_k^\alpha = 0 \tag{12}$$

$$x_k(\rho^\alpha) = [U_k^{-1}(\lambda_k^\alpha)]_{m_k}^{M_k} \tag{13}$$

In this case, it $\rho^{\alpha *}$ is the optimal value of the written dual function, $x_k(\rho^{\alpha *})$ which can also be considered as the optimal of the initial function. The gradient projection method has been used to solve the dual problem. In this method, the value ρ^α is set in the opposite direction of the gradient $\nabla D(\rho^\alpha)$.

$$\rho_l^\alpha(t+1) = \left[\rho_l^\alpha(t) - \psi \frac{\partial D(\rho^\alpha(t))}{\partial \rho_l^\alpha} \right]^+ \tag{14}$$

$$\frac{\partial D(\rho^{\alpha}(t))}{\partial \rho_l^{\alpha}} = c_l - \sum_{k \in K(l)} x_k(\rho^{\alpha}) \quad (15)$$

$$\rho_l^{\alpha}(t+1) = \left[\rho_l^{\alpha}(t) - \psi(\sum_{k \in K(l)} x_k\left(\rho^{\alpha}(t)\right) - c_l)\right]^{+} \quad (16)$$

ψ in Equation 16 represents the value of the convergence hop. This parameter is an important factor in the convergence speed of the proposed algorithm. The selection of the value of this parameter in the implementation is based on the values mentioned in the reference [28]. When the sum of the costs of using link l is more than the bandwidth of link l, in other words, when the value of the relationship $[\sum_{k \in K(l)} x_k(\rho^{\alpha}(t)) - c_l]$ is positive, the link cost will increase in the next hop (line 2 in the pseudo-code of the proposed algorithm). The pseudo-code of the proposed algorithm for the flow control problem in the form of an iterative algorithm has been shown in Figure7.

4- Numerical Simulation of the Proposed Algorithm

Implementation of the proposed algorithm on the WiNoC with 36 processing cores (four wireless routers (yellow nodes) and 32 wireless routers (0 blue nodes)) and 64 links (four wireless links (dotted links) and 60 wired links) organized in a 6 × 6 mesh structure (Figure 6) have been investigated using the CVX tool [29]. The bandwidth of wired links at 1 Gbps and wireless channels at 2 Gbps have been considered. The traffic pattern generated by the cores is uniform. In this pattern, each core can send or receive packets from all cores in the network.

This implementation also shows the flow rate convergence of cores to an equilibrium point. The transmitter of each flow also prevents congestion in the network by adjusting the rate of the transmitted flow, while using the available bandwidth of the links more effectively. Figure 8 has shown that the proposed algorithm with the step size length $\psi = 3/(1+t)$ has converged the flow rates of all cores to the optimum point after 60 iterations. Baseline routers equipped with wireless antennas can have a higher flow rate at the beginning of work. However, after the 30th iteration, the flow rate of the desired nodes is also decreased because all packets forward to distant destinations must be sent through wireless routers. This problem will cause congestion at wireless routers. Therefore, the flow rate of cores has decreased significantly compared to the initial time. Also, with the step size length $\psi = 1/(1+t)$, the proposed algorithm can converge the flow rate of cores to the final equilibrium point after 91 iterations (Figure 9). The flow rates at 50, 100, and 150 iterations for two scenarios are shown in the proposed algorithm.

Figure 10 shows that $\psi = 1/(1+t)$ the rate of flows overlap approximately after the 93rd iteration and remain unchanged. While for $\psi = 3/(1+t)$, rates are unchanged from the 50th iteration. This problem will prove the rapid convergence rate of the proposed algorithm (Figure 11).

Figure 12 shows the network throughput under uniform and Bit-Complement traffic patterns. As it can be seen, the use of the proposed method has been able to improve the operational capacity of the network. Controlling and adjusting the flow rate of each core according to the capacity of each link can reduce the congestion traffic at each link. Hence, congestion in routers is reduced and the throughput of the network is increased.

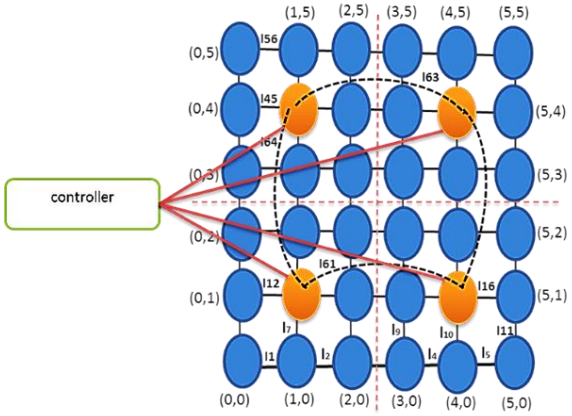


Fig. 6. Network-on-chip with 36 nodes in mesh networks

Initialization:

1. *Initialize C_l of all links.*
2. *Set link price vector to zero.*

Main loop

Do Until $\max \left| x_k^{(t+1)} - x_k^{(t)} \right| \prec 0$

Link Price Update:

1) Received rates $x_k(t)$ *from all sources* $k \in K(l)$

2) Update price by

$$\rho_l^{\alpha}(t+1) = \left[\rho_l^{\alpha}(t) + \psi(\sum_{k \in K(l)} x_k(\rho^{\alpha}(t)) - c_l)\right]^{+}$$

3) Send $\rho_l^{\alpha}(t+1)$ *to all sources* $k \in K(l)$

Rate Adaption:

Received link prices from all links $l \in L(k)$

1) Calculate $\lambda_k^{\alpha}(t) = \sum_{l \in L(k)} A_{lk} \rho_l^{\alpha}$

2) Adjust rate by

$$x_k(\rho^{\alpha}) = \left[U_k^{'-1}(\lambda_k^{\alpha})\right]_{m_k}^{M_k}$$

3) Send $x_k(t+1)$ *to all links.*

Fig. 7. Pseudo code of the proposed algorithm to control the flow rate of WiNoC

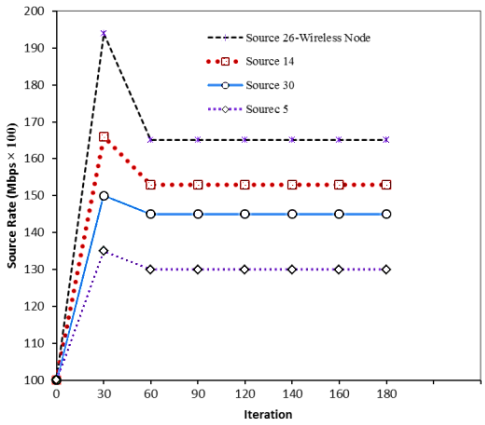


Fig. 8. Core flow rate with uniform hop length of $\psi = 3/(1+t)$ and uniform traffic pattern

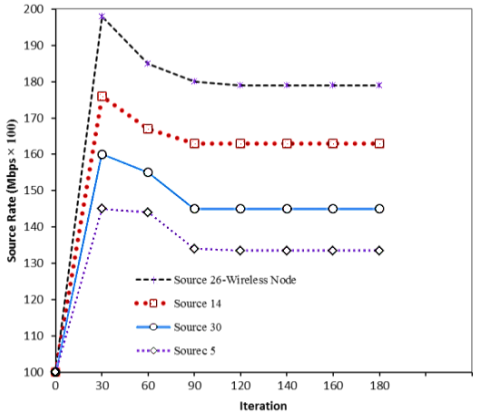


Fig. 9. Core flow rate with uniform hop length of $\psi = 3/(1+t)$ and uniform traffic pattern

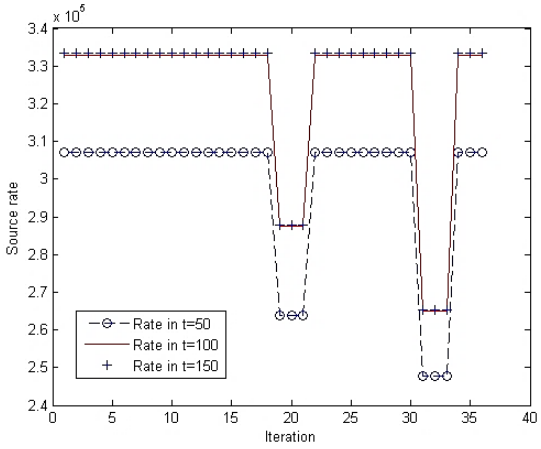


Fig. 10. The convergence of the rates at the iterations 50, 100, and 150 in a uniform, hop-by-hop traffic pattern

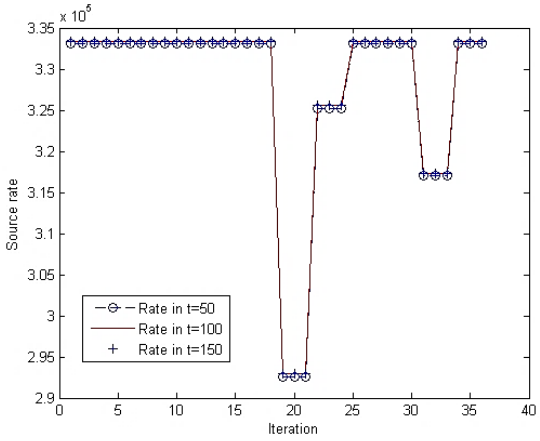


Fig. 11. The convergence of the rates at the iterations 50, 100, and 150 with Bit-Complement traffic pattern and hop length of $\psi = 3/(1+t)$

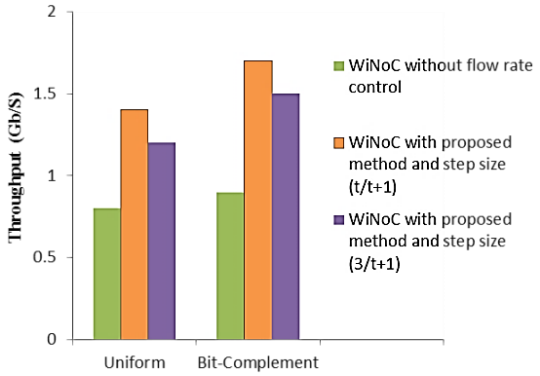


Fig. 12. Network throughput in uniform and Bit-Complement traffic patterns

5- Conclusion

In recent years, the development of the integrated circuit industry will promise to increase the number of in-chip processing cores. In multicore chips, NoC is an effective structure for exchanging data packets among cores. However, due to the communication delay between long-distance cores, congestion is occurring. Therefore, providing an interconnection network with less delay and power consumption will play a key role in improving the performance of multicore systems.

Using the shared wireless links with higher bandwidth on NoCs can reduce the data packet transmission delay. However, the wireless routers on these chips are converted to hot spots because of increasing the demand for data packets forwarding from one area to another through wireless links. Therefore, improving the quality of service and efficient use of resources in WiNoCs cannot be obtained without the presence of mechanisms for flow control or congestion control. In this paper, the flow rate control mechanism in a wireless NoC is introduced as a

utility maximization problem. The solution to this problem is done by writing a dual problem of the primary problem and using the gradient projection method by considering a utility function. Hence, an iterative algorithm is proposed to regulate and control the flow rate of cores. The simulation results of the proposed algorithm under traffic patterns show that the proposed algorithm can regulate the flow rate of processing cores and the network throughput at an acceptable speed and level. As well, the convergence speed of the proposed method under the amount of step size is shown in the proposed algorithm.

References

[1] International Technology Roadmap for Semiconductors. Available: http://www.itrs.net

[2] A. Ben Achballah, S. Ben Othman and S. Ben Saoud, "Problems and challenges of emerging technology networks−on−chip: A review", Microprocessors and Microsystems, vol 53, pp. 1-20, 2017.

[3] J. Lin Jr, H.-T. Wu, Y. Su, L. Gao, A. Sugavanam, and J. E. Brewer, "Communication using antennas fabricated in silicon integrated circuits," Solid-State Circuits, IEEE Journal of, vol. 42, pp. 1678-1687, 2007.

[4] S. Deb, A. Ganguly, P. P. Pande, B. Belzer, and D. Heo, "Wireless NoC as interconnection backbone for multicore chips: Promises and challenges," Emerging and Selected Topics in Circuits and Systems, IEEE Journal on, vol. 2, pp. 228-239, 2012.

[5] D. Zhao and Y. Wang, "SD-MAC: Design and synthesis of a hardware-efficient collision-free QoS-aware MAC protocol for wireless network-on-chip," Computers, IEEE Transactions on, vol. 57, pp. 1230-1245, 2008.

[6] S.-B. Lee, S.-W. Tam, I. Pefkianakis, S. Lu, M. F. Chang, and C. Guo, "A scalable micro wireless interconnect structure for CMPs," in Proceedings of the 15th annual international conference on Mobile computing and networking, pp. 217-228, 2009.

[7] K. Kempa, J. Rybczynski, Z. Huang, K. Gregorczyk, A. Vidan and B. Kimball, "Carbon nanotubes as optical antennae," Advanced Materials, vol. 19, pp. 421-426, 2007.

[8] Abadal, S., Llatser, I., Mestres, A., Solé-Pareta, J., Alarcón, E. and Cabellos-Aparicio, "Fundamentals of Graphene-Enabled Wireless On-Chip Networking," In Modeling, Methodologies, and Tools for Molecular and Nano-scale Communications, pp. 293-317, 2017.

[9] K. Chang, S. Deb, A. Ganguly, X. Yu, S. P. Sa,h and P. P. Pande, "Performance evaluation and design trade-offs for wireless network-on-chip architectures," ACM Journal on Emerging Technologies in Computing Systems (JETC), vol. 8, p. 23, 2012.

[10] M. Palesi, M. Collotta, A. Mineo, and V. Catania, "An Efficient Radio Access Control Mechanism for Wireless Network-On-Chip Architectures," Journal of Low Power Electronics and Applications, vol. 5, pp. 38-56, 2015.

[11] Dehghani, A. and Jamshidi, K.," A fault-tolerant hierarchical hybrid mesh-based wireless network-on-chip architecture for multicore platforms," The Journal of Supercomputing, vol. 71, no. 8, pp. 3116-3148, 2015.

[12] U. Y. Ogras and R. Marculescu, "It's a small world after all": NoC performance optimization via long-range link insertion," Very Large-Scale Integration (VLSI) Systems, IEEE Transactions on, vol. 14, pp. 693-706, 2006.

[13] Hu, W. H., Wang, C., and Bagherzadeh, N," Design and analysis of a mesh-based wireless network-on-chip," The Journal of Supercomputing, vol. 71, no. 8, pp. 2830-2846, 2015.

[14] Ogras, U. Y. and Marculescu, R," Analysis and optimization of prediction-based flow control in networks-on-chip," In Modeling, Analysis and Optimization of Network-on-Chip Communication Architectures, pp. 105-133, 2013.

[15] A. Pullini, F. Angiolini, D. Bertozzi, and L. Benini, "Fault tolerance overhead in network-on-chip flow control schemes," in Integrated Circuits and Systems Design, 18th Symposium on, pp. 224-229, 2005.

[16] J. W. van den Brand, C. Ciordas, K. Goossens, and T. Basten, "Congestion-controlled best-effort communication for networks-on-chip," in Proceedings of the conference on Design, automation, and test in Europe, pp. 948-953, 2007.

[17] A. Rezaei, M. Daneshtalab and D. Zhao," CAP-W: Congestion-Aware Platform for Wireless-based Network-on-Chip in Many-Core Era," Microprocessors and Microsystems, vol. 52, pp. 23-33, 2017.

[18] T. Marescaux, A. Rangevall, V. Nollet, A. Bartic, and H. Corporaal, "Distributed Congestion Control for Packet-Switched Networks on Chip," in ParCo, pp. 761-768, 2005.

[19] F. Jafari, M. S. Talebi, A. Khonsari, and M. Yaghmaee, "A Novel Congestion Control Scheme in Network-on-Chip Based on Best Effort Delay-Sum Optimization," in Parallel Architectures, Algorithms, and Networks, International Symposium on, pp. 191-196, 2008.

[20] Talebi, M. S., Jafari, F., Khonsari, A. and Yaghmaee, M. H," Proportionally fair flow control mechanism for best-effort traffic in network-on-chip architectures", International Journal of Parallel, Emergent, and Distributed Systems, vol. 25, no. 4, pp. 345-362, 2010.

[21] Durand, Y., Bernard, C. and Clermidy, F," Distributed Dynamic Rate Adaptation on a Network on Chip with Traffic Distortion," In Embedded Multicore/Many-core Systems-on-Chip (MCSoC), pp. 225-232, 2016.

[22] Y. Wang and D. Zhao, "Distributed flow control and buffer management for Wireless Network-on-Chip," in Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on, pp. 1353-1356, 2009.

[23] C. Wang, W.-H. Hu, and N. Bagherzadeh, "A wireless network-on-chip design for multicore platforms," in Parallel, Distributed and Network-Based Processing (PDP), 19th Euromicro International Conference on, pp. 409-416, 2011.

[24] D. Zhao and R. Wu, "Overlaid mesh topology design and deadlock-free routing in wireless network-on-chip," in Networks on Chip (NoCs), Sixth IEEE/ACM International Symposium on, pp. 27-34, 2012.

[25] S. H. Low and D. E. Lapsley, "Optimization flow control—I: basic algorithm and convergence," IEEE/ACM Transactions on Networking (TON), vol. 7, pp. 861-874, 1999.

[26] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," Journal of the Operational Research Society, pp. 237-252, 1998.

[27] S. Boyd and L. Vandenberghe, Convex optimization: Cambridge university press, 2004.
[28] S. Boyd, "Convex Optimization II Lecture Notes," Ed: Stanford University, 2006.
[29] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," ed, 2008.
[30] Mohtavipour, S. M., Mollajafari, M. and Naseri, A, ‘‘a novel packet exchanging strategy for preventing HoL-blocking in fat trees,’’ Cluster Computing, pp: 1-22, 2019.
[31] Mortazavi, H., Akbar, R., Safaei, F. and Rezaei. A., ‘‘a fault-tolerant and congestion-aware architecture for wireless networks-on-chip,’’ Wireless Networks, vol. 25, no. 6, pp. 3675-3687, 2019.
[32] Yazdanpanah, F., AfsharMazayejani, R., Alaei, M., Rezaei, A. and Daneshtalab, M., “An energy-efficient partition-based XYZ-planar routing algorithm for a wireless network-on-chip”, The Journal of Supercomputing, vol. 75, no. 2, pp. 837–861, 2019.
[33] Rezaei, A., Daneshtalab, M., Zhao and D. CAP-W, “Congestionaware platform for wireless-based network-on-chip in many-core era”, Microprocessors and Microsystems, vol. 52, pp. 23–33, 2017.
[34] S. M. Mamaghani and M. A. J. Jamali, ‘‘A load-balanced congestion-aware routing algorithm based on time interval in wireless network-on-chip,’’ Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 7, pp. 2869-2882, 2019.

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 57-64
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Study and Realization of an Alarm System by Coded Laser Barrier Analyzed by the Wavelet Transform

Brahim Meriane [1,2*], Salah Rahmouni [2], Issam Tifouti [2]

[1]. University of Batna 2 (UB2). Algeria, Advanced Electronics Laboratory
[2]. Higher School of Technological Education Skikda ENSET. Laboratory (LPCBM) Algeria.

Received: 30Aug 2021/ Revised: 20 Mar 2022/ Accepted: 24Apr 2022

Abstract

This article introduces the study and realization of the laser barrier alarm system, after the laser is obtained by an electronic card, the wireless control system is connected to the control room to announce the application in real time, and the laser is used in many applications fields, from industry to medicine, in this article on the basis of Industrial applications such as laser barrier. It uses an alarm system to detect and deter intruders. Basic security includes protecting the perimeter of a military base or a safety distance in unsafe locations or near a government location. The first stage secures surrounding access points such as doors and windows; The second stage consists of internal detection with motion detectors that monitor movements, In this article, we adopt the embodiment of a coded laser barrier that is transmitted between two units, processes the signal, compares the agreed conditions, and to be high accuracy, we suggest using wavelet transmission to process the received signal and find out the frequencies that achieve alarm activation considering that the transmitted signal They are pulses, but after analysis with a proposed algorithm, we can separate the unwanted frequencies generated by the differential vibrations in order to arrive at a practically efficient system.

Keywords : Wavelet Transform; Lasers Sources ; Radio Frequency; Laser Coded Barriers; Alarm System.

1- Introduction

In recent years, warning and protection systems have been developed in several fields, similar to the military field, where lasers have been used to detect any attempt to penetrate the wall of military barracks. In addition to the existing technologies in this field, we are working on the use of coded lasers, which means that we send very limited pulses in frequency and periodic ratio, as well as in the number of pulses during a pre-agreed period of time[1].
Laser pulses can be obtained through an electronic circuit with analog processing, and to eliminate any noise in the receiving circuit, we filter the signal using wavelet transformation, thanks to which we get high accuracy and an effective system that works in real time[2].
In addition to activating the alarm, this system can also work to send information via radio waves to the control room so that the leadership can make decisions at the same time[2,3].

2- Photovoltaic Barriers

Photovoltaic barriers are optical or electronic systems consisting of a sensor (receiver) and a light source (emitter). The light source can be an ordinary lamp, an infrared emitter (for example, a pulse), LEDs or a laser emitter [2].

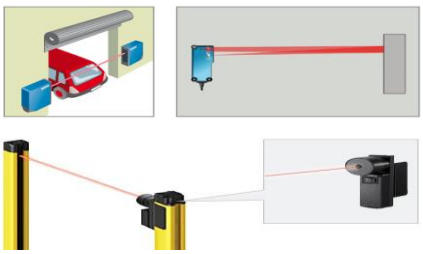


Fig.1 Photovoltaic barriers

✉ **Brahim Meriane**
tlcom_brahim@yahoo.fr

3- Laser Barrier Application

Single barriers consist of a separate interacting transmitter and receiver. Reflex barriers and detectors combine sensor and light source in a single box. In reflex barriers, the emitted light beam is returned by a reflector (prism, reflective sheet) to the receiver.. [4] ,Light barriers register an interruption in the light beam and convert the information. If an object passes through the beam of an optoelectronic barrier, the sensor generates a predefined electrical output signal. It triggers an alarm. Detectors send a very fine infrared beam and react to the reflection of light from an object. The maximum detection distance depends largely on the reflectance rate, shape, color and surface quality of the material[5,6,13].

4- System Structure

The corresponding figure shows the stages of transmitting and processing information that determines with high accuracy all the electronic circuits on which this project depends, as it consists of a laser transmitter encoded between two transmitting and receiving units, the processing stage, and the activation of the alarm with the radio wave communication system.

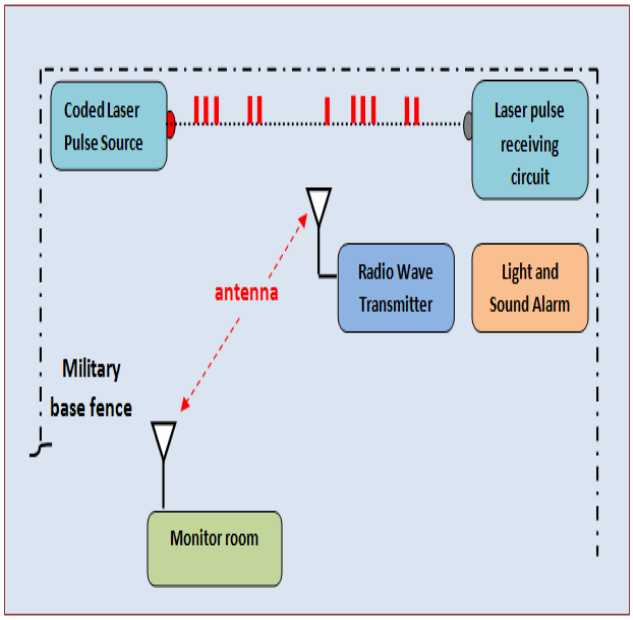


Fig.2 Configuration of the whole system

5- Laser Transmitter Circuit

The corresponding circuit shows the electronic card responsible for producing the laser pulses, with the possibility of changing the frequency and the periodic ratio[7].

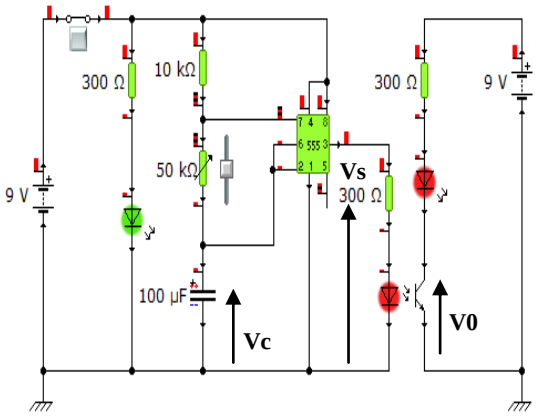


Fig.3 Lasertransmission circuit

Circuit Diagram simulated in Crocodile Technology 607 has shown in Fig-4

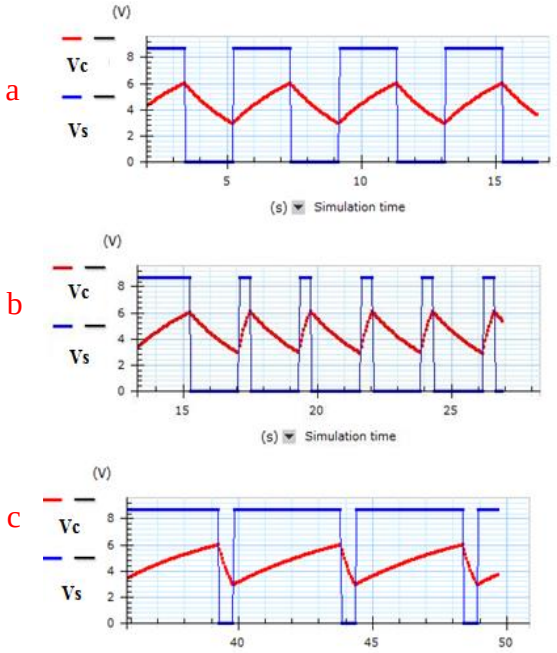
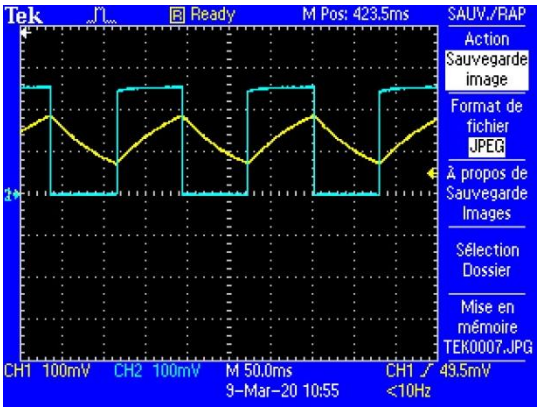


Fig.4 Simulation results of the laser transmitter circuit
a . square electric signal , b . pulse signal ,
c . rectangularsignal

Applied results of the transmission circuit obtained using an oscilloscope

a



b

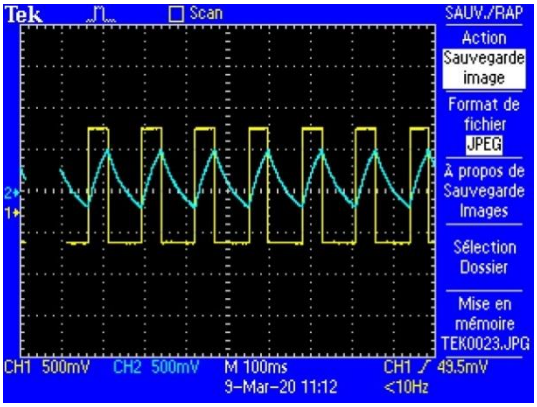


Fig.5The real results of the laser transmitter circuit
a . square electric signal, b . pulse signal

6- Laser Receiver Circuit

Comparator is one of the most important components of analog integrated circuits [14] ,The receiving circuit relies on a practical amplifier that compares the voltages E+ and E-

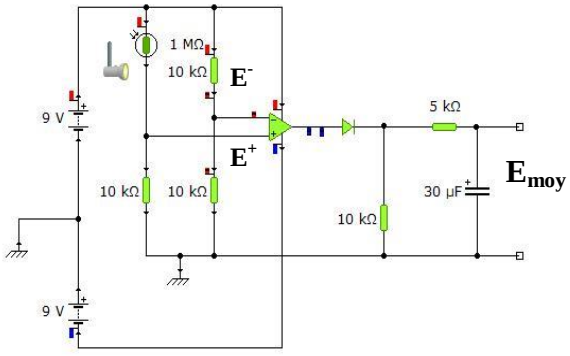


Fig.6 Laser receiving circuit using comparator LM741

After receiving and processing the laser beam, we get the following signal

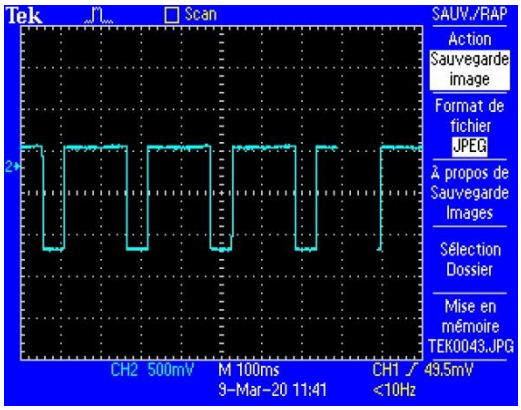


Fig.7 Transmission results by changing the periodic ratio

The signal obtained is either square, rectangular or pulsed, depending on the transmission signal, and accordingly, the average value can be calculated as follows:

$$\overline{E} = \frac{1}{T}\int_0^T E(t).dt \text{.....(1)}$$

$$\overline{E} = \frac{1}{T}\int_0^{t1} E(t).dt + 0 \text{.....(2)}$$

$$\overline{E} = \frac{1}{T} E.t_1 \text{.....(3)}$$

Knowing that the mean value is proportional to the duration t1 then

$$\overline{E} = \eta.E \text{.....(4)}$$

η : The Cyclic Report

The following circuit is to compare the average value of the main signal and the reference voltage, to increase the accuracy by adding another practical amplifier and some electronic components

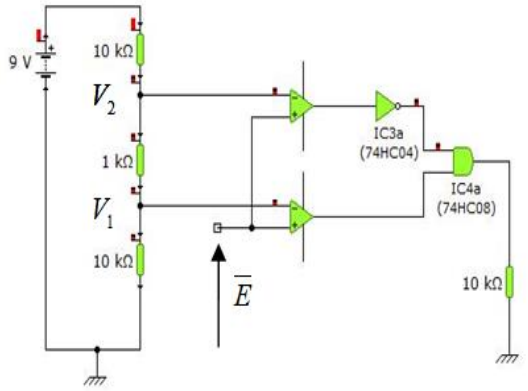


Fig.8 Comparison circuit with two amplifiers LM 741

$$\overline{E} = 2.5V \quad . V1 = 2.3V \quad . V2 = 2.8V$$

So $V_1 \prec \overline{E} \prec V_2$ for this condition the alarm system in the off state because no laser beam cut between the two cards (transmission & reception).

7- Processing Circuit

The main processing circuit consists of the following electronic components

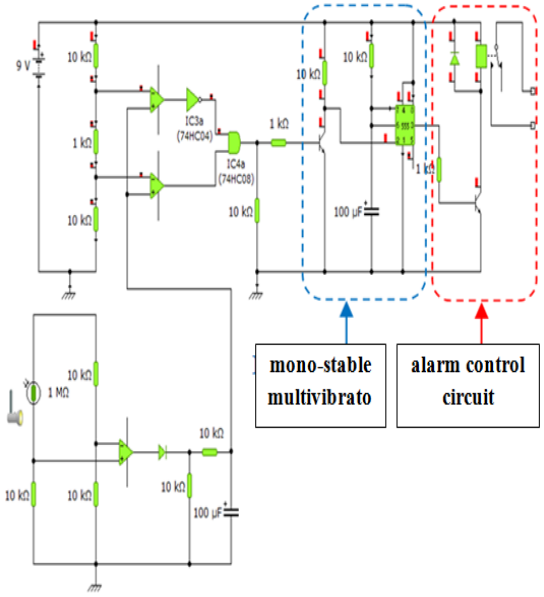


Fig.9 Electronic circuit of the project

The corresponding figure represents the printed circuit of the project using ExpressPCB software and electronic components CMS

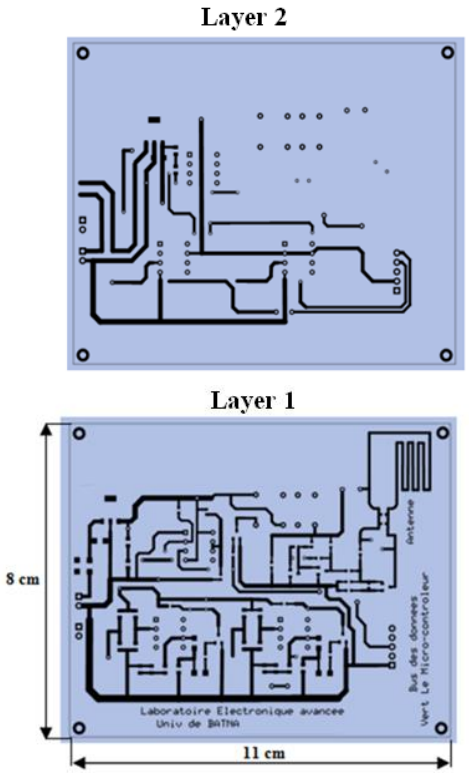


Fig.10 The mother card for the project

The following figure shows the real picture of the project with the processing circuit and data transmission using radio waves.

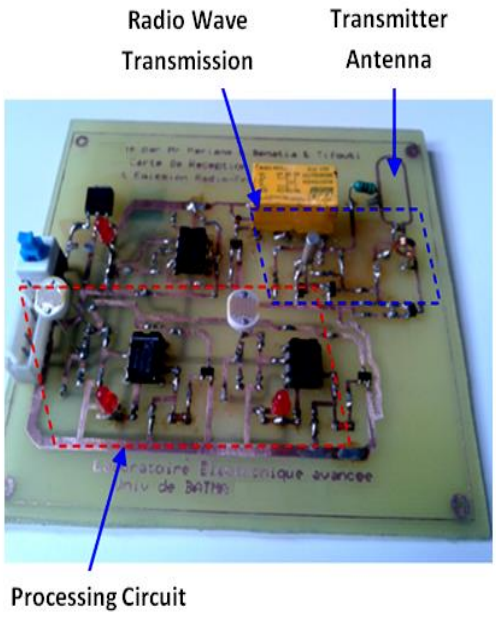


Fig.11 The final electronic card for the project

8- Experimental Results

The figures above show the transmit and receive signal with duty cycle approximately 25% & 75% respectively

Emission signal

Reception signal +/- 12

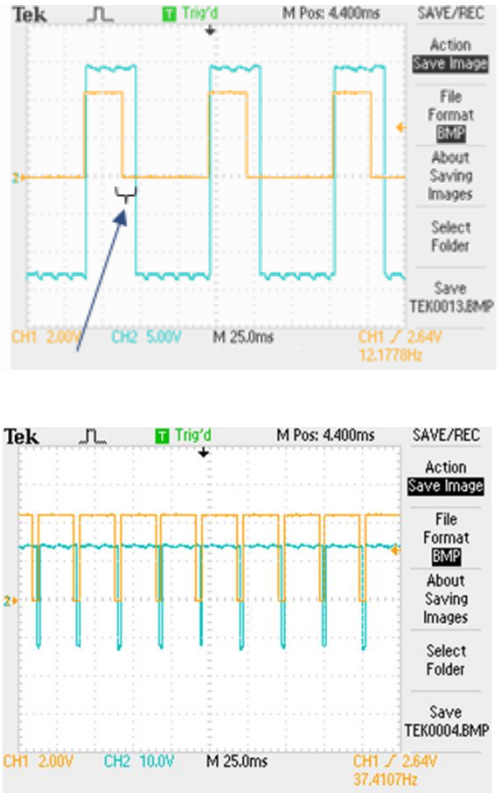


Fig.12 Experimental results

9- Analysis of the Reception Signal by Wavelet Transformation

In recent years, it has widely adopted wavelet transform (WT) and various signal analysis measures in the TF field, and can guarantee the temporal and spectral accuracy in the entire frequency range, as WT has been used in many applications, such as signal deconvolution, imaging processing, noise removal and speed improvement sound classification, etc. [8]. Chu and Kim applied the Morlett wavelet transform to analyze the effect of noise.

Wavelets have two important properties: firstly, the scaling factor, and secondly, the transformation and the relationship between them roughly agrees to the scaling process. Compressed waves are used. When high-bandwidth waves are extended, they correspond to low-frequency signals [9], in low-bandwidths, they correspond to rapidly changing physical signals and magnitudes consisting of high frequencies. In contrast to other theories such as (Fourier transforms, etc.) used in signal processing, waves allow signals to be analyzed in both frequency and time domains. There are two types of wavelet transfers: continuous and discrete existent transports. Both transformations are continuous in time (analog), and with their help analog signals can be represented [10].

10- General Theory of CWT

In this work, we only touched upon some of the basic equations, definitions and concepts of wavelet transform, and a more rigorous mathematical treatment of this topic can be found in [10,11]. The time continuous wavelet transformof f(t) is defined as:

$$CWT_{\psi} f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t).\overset{*}{\psi}\left(\frac{t-b}{a}\right).dt \text{.....(5)}$$

Here $a,b \in R, \quad a \neq 0$ and they are dilating and translating coefficients, respectively. This multiplication of |a|-1/2is for energy normalization purposes so that the transformed signal will have the same energy at every scale. The analysis function Ψ(t), the so-called mother wavelet has to satisfy that it has a zero net area, which suggest that the transformation kernel of the wavelet transform is a compactly support function[10].

We can say that a disadvantage of CWT is that the representation of the signal is often redundant, because a and b are continuous over R (the real number).As the original signal can be completely reconstructed by a model copy of Wf (b, a). Usually, we try Wf (b, a) in a binary network i.e., a = 2-mandb = n2-mm, n∈Z+. Substituting the last one into

$$DWT_{\psi} f(a,b) = \int_{-\infty}^{+\infty} f(t).\overset{*}{\psi}(t).dt \text{.....(6)}$$

where Ψm,n(t) = 2-mΨ(2mt-n) is the dilated and translated version of the mother wavelet Ψ(t)[10,11].

11- Temporal and Spectral Resolutions in the CWT

Resolutions in the time and frequency domains are critical for evaluation of performance of different wavelets. The temporal resolution in the time domain σ_t and the spectral resolution in the frequency domain σ_w of the CWT can be defined as[5]:

$$\sigma_t^2(\gamma) = \int_{-\infty}^{+\infty} (t-u\gamma)^2 \left|\phi_\gamma(t)\right|^2 dt \qquad .....(7)$$

DWT is a mathematical tool for decomposing data in a top-down manner. DWT represents a function in terms of a rough overall form, and a wide range of details. Despite of the requirement and type of function i.e., signals images etc. DWT offers sublime methodology and technique for representing the amount of detail present.

Wavelets perform and offer scale based analysis for a given data. A wide range of applications and usage has been found for these wavelets including signal processing, mathematics and numerical analysis and, for its better performance in signals and image processing it is considered an alternative to Fast Fourier Transform as DWT provide time frequency representation When there is a need for processing and analyzing non stationary tool, DWT can be used .Study shows that discrete wavelets transform have a high performance in speech signal processing so far [9,12].

Computer-assisted experiment (CAM) is not fundamentally different from experimentation as it has been traditionally performed using different measuring instruments and laboratory equipment, but the integration of the new algorithms in the signal processing of the laser pulses brings many advantages. The data acquisition process can be automated, and the measurement results can be easily saved and manipulated by various software tools. In addition, the presentation of the results in a graphical form is greatly simplified, which facilitates the analysis and use of wavelet transform on the obtained signal.

12- Acquisition of Reception Signal with Cooledit

CoolEdit program is used to record electrical signals that are proportional to the physical quantities to be processed and will be recorded in a one-dimensional matrix at a sampling frequency of 64 kHz using a 16 transducer called the sampling period.

$$T_e = \frac{1}{F_e} = \frac{1}{64\times10^3} = 0.015\times10^{-3}\,s \ldots\ldots (8)$$

The following design represents how to record the electrical signal from the receiving circuit using computer-assisted experiments.

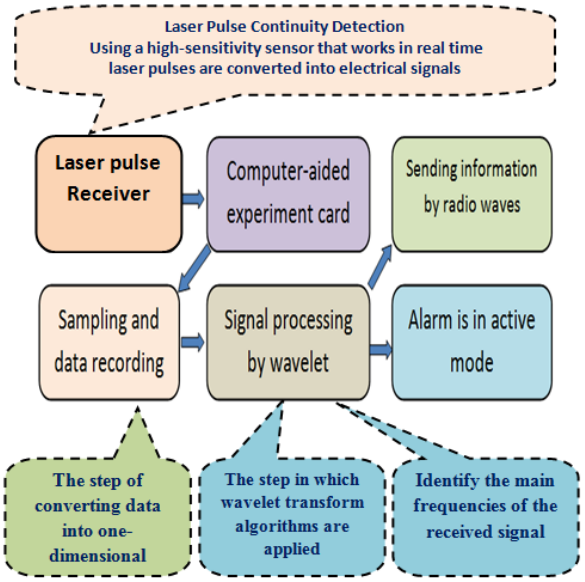


Fig.13 Diagram of the proposed method

The signal obtained is considered a non-stationary signal and it is also made up of several signals which cannot be recorded using the oscilloscope because it is technically unable to track and oscillate instantaneous signals of very high frequency. Response time, although it is very infinite, and also takes a slow motion until it stabilizes, and this is called differential vibration.

The molar figure shows the difference between the signal recorded by the cathode oscilloscope and the one recorded by Computer-assisted experiment.

There is no differential vibration

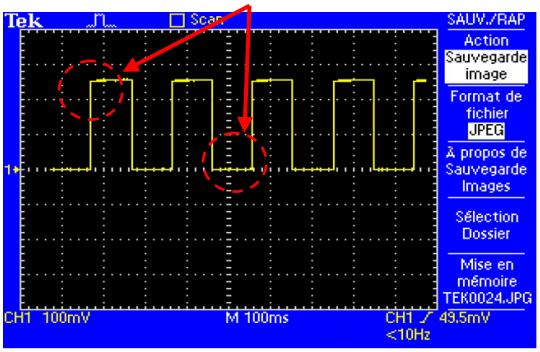


Fig.14 The results of the transmitter circuit by the oscilloscope

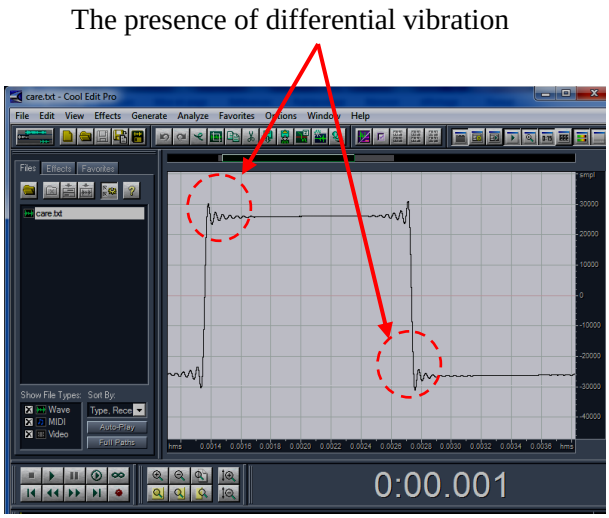


Fig.15 Real results of the transmitter circuit by CoolEdit

The fluctuation of laser signals is a very important measure. CWT is often applied to detect the singularity numerical of a transient signal at each stage of a simulation, standard procedures for wavelet computation have been used coefficients (in the case of continuous as well as discrete wavelet transitions), which are integral parts of the MATLAB program. For these results we used the CWT function.

After processing the stored matrix, we get the following figure, after applying the wavelet transform with the selection of the Haar wavelet algorithm to analyze the signal in several levels, where we can know the basic frequencies of the received laser signal.

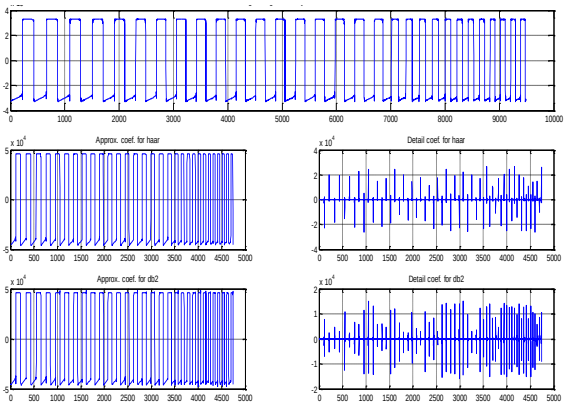


Fig.16Simulation results of wavelet transformation on the receiving signal at an increased frequency

The following figure shows how to deconstruct the signal obtained in the receiving circuit, showing the frequency resulting from the differential vibration, which we would not have obtained without applying the wavelet theory.

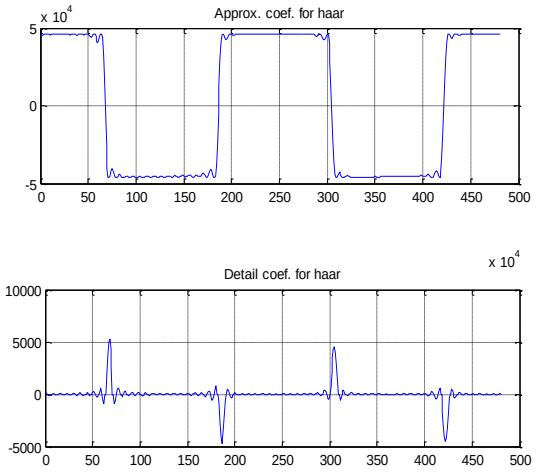


Fig.17 the Haar wavelet transform on the received signal

13- Conclusions

In this work, we relied on processing the unstable signals sent between the transmitter and the receiver, which are coded laser pulses that control a sophisticated alarm system that can be used in the military field. In order for this system to be effective, we decided to analyze the received signal by converting waves, which have proven successful in several areas. In practice, whenever we can decompose the received signal using wavelet transformation, the results are perfect. It can be said that the transmitted pulses consist of several signals with different frequencies, and this leads to the lack of dependence on this signal in controlling any of them. The system, especially if it requires high accuracy or efficiency in performance such as controlling the speed of the DC motor or transmitting digital information by wireless communication, especially if it is high frequency. On this basis, we proposed an algorithm to disassemble and filter the received signal in order to have effective and satisfactory results.

References

[1] Muhammad Asyraf Zainal Arifin “Automation Security System with Laser Lights Alarm on Web Pages and Mobile Apps ” 978-1-5386-8546-4/19/$31.00 ©2019 IEEE.

[2] A. S. Life. 2019’s Best Home Security Systems. Available:https://www.asecurelife.com/best-home-security-system/,Accessed on 7/1/2019.

[3] A. Tseloni, R. Thompson, L. Grove, N. Tilley, and G. Farrell, "The effectiveness of burglary security devices," Security Journal, vol. 30, pp. 646-664, 2017.

[4] Ashis Rai ‘’ Low Cost Laser Light Security System in Smart Home ‘’ 978-1-7281-2443-8/19/$31.00 ©2019 IEEE.

[5] Arianne B ‘’ Design and Implementation of an Arduino-Based Security System Using Laser Light ‘’ LPU-Laguna Journal of Engineering and Computer Studies Vol. 4 No. 2 October 2019.
[6] S.Rahmouni, L. Zighed, S.Chaguetmi, M. Daoudi, M. Khelifa, M. Karyaoui, R. Chtourou. ‘’ Correlation between photoluminescence and ellipsometric measurements of porous silicon layers‘’.(Optoelectron. Adv. Mat.) . Vol.12, 9-10 ; P 553-558. (2018).
[7] Md. Moyeed Abrar ‘’ Design and Implementation of AstableMultivibrator using 555 Timer ‘’ IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) , Volume 12, Issue 1 Ver. II (Jan. – Feb. 2017), PP 22-29.
[8] J. Kim “Time-frequency characterization of hand-transmitted, impulsive vibrations using analytic wavelet transform,” Journal of Sound and Vibration, vol.308, pp. 98-111, Nov 2007.
[9] Singh, S.D. Joshi, R.K. Patney, K. Saha, ‘’ The Fourier decomposition method for nonlinear and non-stationary time series analysis ‘’, Proc. R. Soc. A 473 (2017) 2016.0871.
[10] Mohammad RasoulKahrizi "Long-Term Spectral Pseudo-Entropy (LTSPE): A New Robust Feature For Speech Activity Detection"Journal of Information Systems and Telecommunication, Vol. 6, No. 4, October-December 2018.
[11] B Meriane ‘’Denoising and Enhancement Speech Signal Using Wavelet ‘’ Journal of Information Systems and Telecommunication, Vol. 9, No. 1, January-March 2021.
[12] RaziehTorkamani, Sadeghzadeh "Wavelet-based Bayesian Algorithm for Distributed Compressed Sensing", Journal of Information Systems and Telecommunication, Vol. 7, No. 2, April-June 2019.
[13] Ruiqiang Guo, Lu Song "Space Laser Chaotic Security System", 978-1-5386-3573-5/17/$31.00 ©2017 IEEE
[14] Anil Khatak, Manoj Kumar, Sanjeev Dhull ‘’ An Improved CMOS Design of Op-Amp Comparator with Gain Boosting Technique for Data Converter Circuits ‘’ Journal of Low Power Electronics and Applications, 2018, 8, 33

Journal of Information Systems and Telecommunication
Vol.11, No.1, January-March 2023, 65-74
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Foreground-Back ground Segmentation using K-Means Clustering Algorithm and Support Vector Machine

Masoumeh Rezaei[1*], Mansoureh Rezaei[2] , Masoud Rezaei[3]

[1]. Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Zahedan, Iran
[2]. Computer Engineering Department, Faculty of Engineering, Yazd University, Yazd, Iran
[3]. Faculty of Electrical Engineering, K. N. Toosi University of Technology, Tehran, Iran

Received: 10 Jun 2021/ Revised: 04 Apr 2022/ Accepted: 06 May 2022

Abstract

Foreground-background image segmentation has been an important research problem. It is one of the main tasks in the field of computer vision whose purpose is detecting variations in image sequences. It provides candidate objects for further attentional selection, e.g., in video surveillance. In this paper, we introduce an automatic and efficient Foreground-background segmentation. The proposed method starts with the detection of visually salient image regions with a saliency map that uses Fourier transform and a Gaussian filter. Then, each point in the maps classifies as salient or non-salient using a binary threshold. Next, a hole filling operator is applied for filling holes in the achieved image, and the area-opening method is used for removing small objects from the image. For better separation of the foreground and background, dilation and erosion operators are also used. Erosion and dilation operators are applied for shrinking and expanding the achieved region. Afterward, the foreground and background samples are achieved. Because the number of these data is large, K-means clustering is used as a sampling technique to restrict computational efforts in the region of interest. K cluster centers for each region are set for training of Support Vector Machine (SVM). SVM, as a powerful binary classifier, is used to segment the interest area from the background. The proposed method is applied on a benchmark dataset consisting of 1000 images and experimental results demonstrate the supremacy of the proposed method to some other foreground-background segmentation methods in terms of ER, VI, GCE, and PRI.

Keywords: Foreground-Background Segmentation; Support vector machine; k-means clustering; saliency map.

1- Introduction

Image segmentation is one of the most significant image processing tasks in image analysis and understanding. The main goal of image segmentation is finding objects of interest from a given image using image characteristics such as color, texture, gray level and, so on. Typically, image segmentation methods can be categorized into six categories, Edge detection-based methods, Histogram thresholding-based methods, Graph-based methods, Region-based methods, Statistical model-based methods, and Machine learning-based methods [1].

Histogram thresholding-based approaches use the assumption that adjacent pixels whose value lies within a certain range belong to the same class. Because of their intuitive properties, simplicity of implementation, and computational speed image, these techniques are widely used [2-6]. Edge detection-based approaches assume that pixel values at the boundary between two regions change quickly. There are some edge detectors such as Canny [7], Prewitt [8], Sobel [9], and so on. The output of edge detectors provides candidates for the region boundaries. These algorithms are only suitable for noise-free and simple images [10,11]. The region-based approaches assume that adjacent pixels in the same region have similar visual characteristics [12-15]. By using these methods, pixels can be grouped into homogeneous regions that might be corresponding to an object.

In the Graph-based methods, an image is considered a weighted graph. Pixels and similarities between them are considered as nodes and edges of the graph, respectively. In these methods, image segmentation is measured as a problem of partitioning this graph into components with minimizing a cost function [1]. Graph cuts as one of the most important graph-based methods was introduced in 2001 [16].

Statistical model-based methods use a statistical model that characterizes pixel values [17, 18]. Machine learning-

✉ **Masoumeh Rezaei**
mrezaei@ece.usb.ac.ir

based methods use machine learning techniques for image segmentation [19-25]. In the last decade, some classification techniques have been successfully used in image segmentation. Especially, SVM as a binary classifier can be used for this purpose. In 2011, the Fast Support Vector Machine (FSVM) as a modified SVM was introduced for image segmentation [26]. User-selected objects and background pixels are used for training in this method. In the same year, Wang et al. applied Fuzzy C-Means-SVM (FCM-SVM) for color image segmentation [27]. In this method, training samples are selected randomly from the FCM clustering results. The drawback of this method is the number of FCM clusters must be set in advance, and the random selection of training samples also affects the performance of the final segmentation.

The saliency-SVM (SSVM) method is a combination of visual saliency detection and SVM classification [28]. In this method, a trimap of the given image is extracted according to the saliency map for estimating the prominent locations of the objects. Positive and negative training sets are automatically selected for SVM training through histogram analysis in trimap. The entire highlighted object is segmented using a trained SVM classifier. In 2018, Sangale and Kadu introduced a real-time Foreground-background segmentation using C-1SVM (Competing 1-class Support Vector Machines) technique [29]. The method first trains local C-1SVMs at every pixel area. Then, it relabels each pixel using learned C-1SVM. In the next step, it performs matting along the foreground boundary and then it applies global optimization.

Tang et al. applied SVM for Foreground Segmentation in video sequences [30]. They introduced a novel feature image and used it in the framework of a support vector machine. In 2020, the SVM method is proposed for identifying two locust species and instars [31]. They used the Grab Cut method and principal component analysis for extracting eight features from 73 features of locusts. However, the proposed Image segmentation and feature extraction of this method are complicated which causes difficulty to achieve fully automatic identification.

In addition to the SVM-based methods, some other methods have been introduced in this field. Dhanachandra et al. used k-means to segment the foreground area from the background. The subtractive clustering method is applied to generate the centroid based on the potential value of the data points. In this method, partial contrast stretching is used for improving the quality of the image and the middle filter is applied to improve the segmented image [32].

In 2021, Chen et al. proposed an improved K-means algorithm for underwater image background segmentation. The method deals with the problem of improper determination of the value of K and minimizes the effect of the initial centroid position of the grayscale image during the quantification of the gray surface of the conventional K-means algorithm [33].

In recent years, researchers have proposed some deep learning methods for the image segmentation. In 2022, Yang et al. proposed a generative adversarial deep network for foreground and background segmentation. The method avoids trivial decompositions by maximizing mutual information between generated images and latent variables [34].

In this paper, we propose a novel and efficient foreground-background segmentation. First, SVM is used for segmenting the interest area from the background. Then, K-means clustering is applied for selecting the training data of SVM. It restricts the computational efforts in the region of interest. However, before applying the K-means algorithm, the first saliency parts are selected using a saliency map and some efficient operators.

The rest of this paper is organized as follows. In Section 2, The SVM and K-Means methods are explained in detail. Section 3 describes the proposed method. Section 4 illustrates the experimental results. Finally, the paper is concluded in Section 5.

2- Primary Concepts

This section provides a detailed description of SVM and K-Means methods as the approaches along with the proposed method.

2-1- Support Vector Machine

SVM was proposed by Vapnik and coworkers [35]. It is a supervised learning method that originated from statistical learning theory. The main idea behind SVM is the separation of the two classes with a hyperplane that maximizes the margin between them. Maximizing margin results in minimizing structural risk. The basis of minimizing structural risk instead of empirical risk is an interesting property of SVM [36]. Thus, SVM outperforms other methods such as neural networks which are based on minimizing empirical risk. Also, its strong generalization reduces the influence of the noise. This method also can be considered a non-linear classification using the kernel trick. The kernel maps their inputs into high-dimensional feature spaces implicitly [37].

Consider the problem of separating the data set of N points $\{x^{(i)}, y^{(i)}\}$ with the input data $x^{(i)} \in R^n$ and the corresponding target $y^{(i)} \in \{-1, +1\}$. In feature space SVM models take the form:

$$f(x) = \omega^T \phi(x^{(i)}) + b \tag{1}$$

where $b \in R$ is a bias term and $\varphi(.): R^n \to R^{n_k}$ is a nonlinear function that maps the input space to a high-dimensional space. The dimension n_k is implicitly defined that it can be infinite-dimensional. SVM optimization problem for the linear separate case is:

$$\min_{w,b} \frac{1}{2}\|\omega\|^2 \qquad (2)$$
$$y^{(i)}(\omega^T \phi(x^{(i)}) + b) \geq 1 \quad i = 1,...,N$$

and for the non-separable case is:

$$\min_{w,b} \frac{1}{2}\|\omega\|^2 + C\sum\nolimits_{n=1}^{N} \xi_i \qquad (3)$$
$$y^{(i)}(\omega^T \phi(x^{(i)}) + b) \geq 1 - \xi_i \quad i = 1,...,N$$

where parameter C is the regularization parameter that controls the tradeoff between the complexity of the model and the training error that needs to be specified a priori. A larger C means assigning a higher penalty to errors. The Lagrangian dual problem for the SVM is simply:

$$L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} K(x^{(i)}, x^{(j)}) \qquad (4)$$
$$s.t. \quad \begin{cases} \sum_{i=1}^{N} \alpha_i y^{(i)} = 0 \\ 0 \leq \alpha_i \leq C \quad i = 1...l \end{cases}$$

where $\alpha_i \in R$ are positive Lagrange multipliers and $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$. Finally, the classification problem is solved using quadratic programming packages. The new data can be classified as follows:

$$\hat{y} = sign(b + \sum_{\alpha_n > 0} \alpha_n y^{(n)} k(x^{(n)}, x)) \qquad (5)$$

$$b = \frac{1}{N_s}(\sum_{\alpha_m > 0} y^{(m)} - \sum_{\alpha_n > 0} \alpha_n y^{(n)} k(x^{(m)}, x^{(n)})) \qquad (6)$$

where N_s is the total number of support vectors.

2-2- K-Means Clustering Algorithm

One of the simplest learning algorithms to solve the clustering problem is K-Means [38]. It partitions a collection of data into k number of disjoint clusters. K-means is an iterative algorithm that has two steps. In the first step, the k centroids are calculated and in the second step, each point is taken to the cluster that has the nearest centroid from the point. It minimizes the sum of distances of each point to its corresponding cluster centroid. The algorithm of k-means clustering is as follows:

Initialize the number of clusters k and centers (e.g., randomly)
For each object,
Calculate the distance d (e. g. Euclidean distance) to each cluster
Assign it to the closest cluster
End
Recalculate the cluster centers positions
Repeat the process until it satisfies the terminal conditions

3- The Proposed Method

The proposed method starts with the detection of visually salient image regions. Saliency detection is the recognition of pixels or regions whose state stands out relative to their neighbors. It defines what is prominent or noticeable. For finding the region of interest in a given image, a saliency map is used. Many researchers have proposed different algorithms for calculating the saliency map. In this paper, the saliency map $SM(x, y)$ is considered as follows [39]:

$$SM(x, y) = g * \left\| F^{-1}[e^{i.p(x,y)}] \right\|^2 \qquad (7)$$

where F^{-1} is the inverse of Fourier Transform, $p(x, y)$ is the phase spectrum of the Fourier transform of a gray level image, and g* represents a 2D Gaussian filter. Using a small standard deviation, as shown in Fig. 1 (c), we hope that the well-defined boundaries of salient objects are achieved. Then, the binarization threshold t is applied for classifying each point on the maps as salient or non-salient (Fig. 1 (d)).

$$T(x, y) = \begin{cases} 0 & if \ SM(x, y) < t \\ 1 & if \ SM(x, y) \geq t \end{cases} \qquad (8)$$

In the next step, the Hole filling algorithm is applied as previously suggested by Soille [40]. Hole filling operators are found in most photo-editing programs, and they are used to fill holes in a given image. Morphologically, a hole is defined as a set of pixels in the background that cannot be reached by filling in the background from the edge of the image. Fig. 1 (e) shows the result after applying the filling holes operation.
The achieved binary image has a small extra spot that needs to be removed. The area-opening method is used for removing objects in the binary image that are too small. By using this method, all connected components with

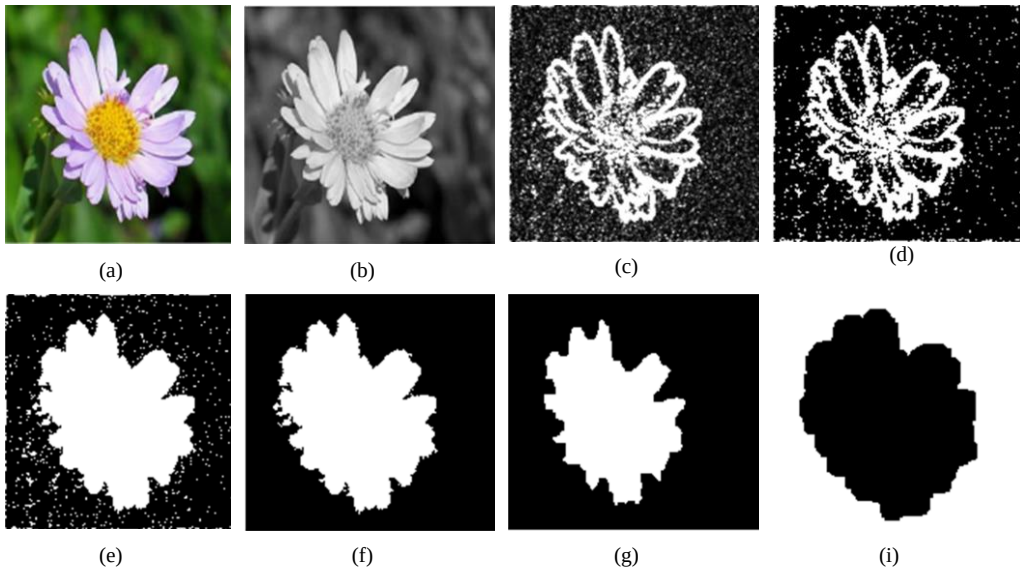


Fig. 1: (a) Original image, (b) Gray level image, (c) saliency maps, (d) After applying binarization threshold, (e) hole filling operator, (f) area opening operator, (g) erosion operator, (h) dilation operator.

fewer than *P* pixels are removed from the image. In this paper, in all experiments, *P* is set to 50. The achieved image is shown in Fig. 1 (f).

For better separation of the foreground and background, dilation and erosion operators are used. Erosion and dilation operators are applied for shrinking and expanding the achieved region (the white region in Fig. 1 (f)), respectively. Two images are realized in this step, one is achieved by applying erosion operator (Fig. 1(g)) and one by applying dilation and then negative operation (Fig. 1(h)). The data in the white region of the first image (F) can be used for foreground samples and the data in the white region of the second image (B) can be used for background samples. Because the number of these data is large, K-means clustering is used as a sampling technique to restrict computational efforts in the region of interest. *K* cluster centers for each region are set for the training of SVM. It causes reducing the complexity of the SVM classifier. In this paper, for each training pixel, six features including *i*, *j* (spatial features), R, G, B, and SM(i,j) are extracted. Finally, all pixels are classified by the trained SVM model.

The algorithm for the proposed method is as follows:

Step 1: Calculate the saliency map for each pixel
Step 2: Classify each point in the maps as salient or non-salient.
Step 3: Fill Holes by applying the Hole filling operator.
Step 4: Remove too small objects with an area-opening operator.
Step 5: Identify F and B regions by applying the erosion and dilation operators.
Step 6: Calculate k cluster centers for each region by applying K-Means
Step 7: Train SVM with achieved data.
Step 8: Classify all pixels by the trained SVM model.

4- Experimental Results

In our experiments, images from a benchmark dataset proposed by Achanta et *al.* [41] are used. The dataset consists of 1000 images. The results are evaluated in terms of PRI, VOI, GCE, and ER. Each of these criteria reveals the capability of segmentation methods from a specific aspect. Given a set of n elements $S=\{o_1,...,o_n\}$ and two partitions of *S*, $X=\{X_1,...,X_r\}$ a partition of *S* into *r* subsets, and $Y=\{Y_1,...,Y_s\}$, a partition of *S* to *s* subsets, the mentioned evaluation metrics are evaluated that described briefly as follows:

I. Probabilistic Rand Index:

The Probabilistic Rand Index (PRI) is introduced as an extension of the Rand Index for evaluating the segmentation approaches [42]. PRI is calculated as follows:

$$PRI(X,Y)=\frac{1}{\binom{n}{2}}\sum_{i,j}C_{ij}p_{ij}+(1-C_{ij})(1-p_{ij}) \quad (9)$$

where C_{ij} is the event that pixels *i* and *j* have the same label and p_{ij} is its probability.

II. Variation of Information

Meila suggested the **Variation of Information** (VI) which measures the amount of information lost and gained in changing from one clustering to another [43]. VI is considered as follows:

$$VI(X,Y) = H(X) + H(Y) - 2*I(X,Y) \quad (10)$$

where *H* and *I* are the entropy and mutual information between two segmentation *X* and *Y*, respectively.

III. Segmentation Error Rate

The segmentation error rate is calculated as:

$$ER = \frac{a+b}{n} * 100\% \quad (11)$$

where a and b are the number of false-segmented image pixels and the number of miss-segmented image pixels, respectively.

IV. Global Consistency Error

The **Global Consistency Error** (GCE) measures the extent to which one segmentation can be viewed as a refinement of the other [44]. For a given element O_i, consider segments that contain O_i in *X* and *Y*. These sets of pixels are denoted by $C(X,O_i)$ and $C(Y,O_i)$, respectively. The GCE can be defined as follows:

$$G(x,y) = \frac{1}{n} \min \left\{ \sum_i \frac{\left| C(X,O_i) / C(Y,O_i) \right|}{C(X,O_i)} , \sum_i \frac{\left| C(X,O_i) / C(Y,O_i) \right|}{C(Y,O_i)} , \right\} \quad (12)$$

In the first experiment, the proposed method was applied to six images. The segmentation results are demonstrated in Fig. 2. It can be found from Fig. 2 that segmentation results are closest to the ground truth segmentation images in all cases. For better evaluation, the mentioned criteria were calculated and reported in Table 1.

The proposed method is also compared with other segmentation methods. The results for those methods are elicited from [28]. The achieved segmented images for five images are shown in Fig. 3. We can see qualitatively that the proposed method outperformed the other methods, and the segmented images achieved by the proposed method are closest to the ground-truth images.

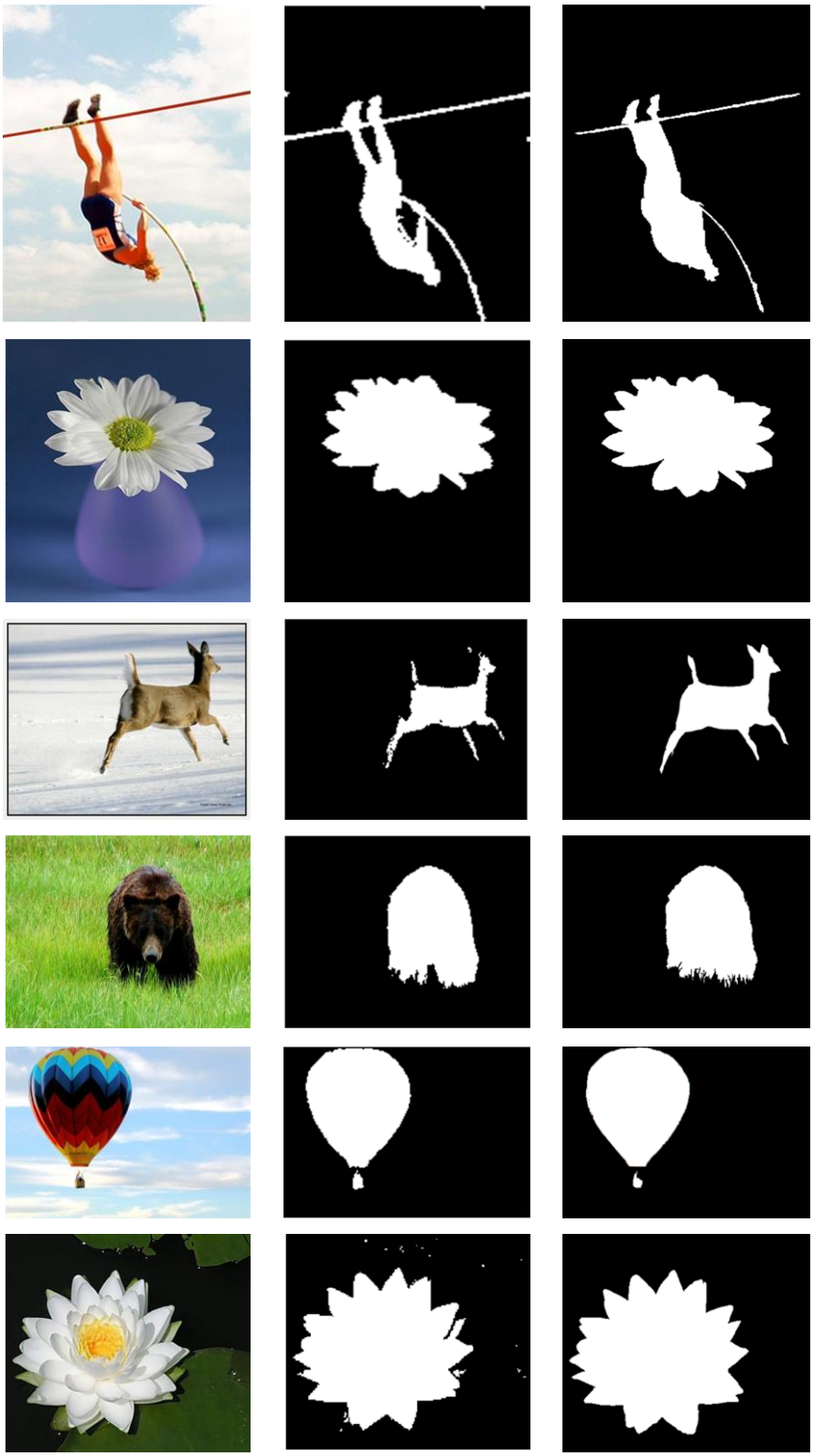
The methods are also evaluated quantitively in terms of ER, VI, GCE and, PRI (Table. 2-5). The results prove the superiority of the proposed method in comparison to other methods. Just in some cases, SSVM could lead to better results. For example, in image #7, SSVM performs better than the proposed method in terms of VI, GCE, and PRI. We can also be found visually in Fig. 1 that the result of SSVM for this image is better than the proposed method.

For a better comparison, the quantitative results on the whole dataset are reported in Table 6. The numbers are average values of ER, GCE, PRI, and VI on the whole dataset. As reported in Table 6, the performance of the proposed method is superior to other segmentation methods in three metrics. But the SSVM and ISVM methods outperformed the proposed method in terms of VI. Unlike pairwise counting comparison criteria, VI is not directly related to the relationships between point pairs. It can be said that VI is based on the relationship between a point and its cluster in each of the two clusters. Note that this is neither a direct advantage nor a disadvantage compared to criteria based on the pair counts.

The proposed method is simple and efficient and the results demonstrate its supremacy in terms of ER, VI, GCE, and PRI.

5- Conclusion

We have developed a novel and efficient foreground-background segmentation using SVM with the K-means clustering method. First, saliency data are automatically selected using a saliency map and some efficient and simple operators, then K-means is applied as a sampling technique to restrict computational efforts in the region of interest. Then, the SVM model is trained using achieved data. We demonstrated the superiority of the proposed approach to some other segmentation methods based on the SVM model in terms of ER, VI, GCE, and PRI on a benchmark dataset consisting of 1000 images. Although the proposed method shows satisfying results in the mentioned criteria, it has one drawback. The proposed method has two free parameters (K and P) which should be allocated with trial and error. In future works, we wish to apply a Convolutional Neural Network to learn saliency features directly.



(a) (b) (c)

Fig. 2: (a) Original images; b) Segmentation results of the proposed method; c) Ground truth segmentations.

Table 1: ER, GCE, PRI and VI of the proposed method for test img1–img6

Image #	VI	PRI	ER	GCE
1	0.2837	0.9537	2.3694	0.0422
2	0.1152	0.9848	0.7681	0.0149
3	0.3679	0.9201	4.1709	0.0554
4	0.1681	0.9774	1.1419	0.0221
5	0.1148	0.9858	0.7143	0.0140
6	0.3232	0.9468	2.7337	0.0510

Table 2: ER of ISVM, FSVM, FCM-SVM, SSVM (Bai and Wang, 2014) and the proposed method for test img7–img11

ER	ISVM	FSVM	FCM-SVM	SSVM	The proposed method
Image #7	9.5	9.7	24.6	5.3	**4.3158**
Image #8	3.9	4.0	4.1	3.4	**0.5932**
Image #9	5.7	7.6	7.3	**0.2**	0.4703
Image #10	8.2	8.7	11.6	**0.4**	0.4727
Image #11	5.9	9.1	5.4	4.7	**1.3275**

Table 3: VI of ISVM, FSVM, FCM-SVM, SSVM (Bai and Wang, 2014) and the proposed method for test img7–img11

VI	ISVM	FSVM	FCM-SVM	SSVM	The proposed method
Image #7	0.29	0.31	0.40	**0.26**	0.4666
Image #8	0.37	0.36	0.36	0.29	**0.0990**
Image #9	0.08	0.13	0.12	**0.04**	0.0782
Image #10	0.19	0.20	0.23	0.14	**0.0771**
Image #11	0.21	0.24	0.20	0.18	**0.1787**

Table 4: GCE of ISVM, FSVM, FCM-SVM, SSVM (Bai and Wang, 2014) and the proposed method for test img7–img11

GCE	ISVM	FSVM	FCM-SVM	SSVM	The proposed method
Image #7	0.07	0.08	0.25	**0.02**	0.0787
Image #8	0.04	0.05	0.07	0.03	**0.0117**
Image #9	0.26	0.30	0.29	0.01	**0.0093**
Image #10	0.20	0.21	0.22	0.02	**0.0092**
Image #11	0.07	0.09	0.06	0.03	**0.0246**

Table 5: PRI of ISVM, FSVM, FCM-SVM, SSVM (Bai and Wang, 2014) and the proposed method for test img7–img11

PRI	ISVM	FSVM	FCM-SVM	SSVM	The proposed method
Image #7	0.89	0.83	0.76	**0.95**	0.9174
Image #8	0.93	0.93	0.94	0.95	**0.9882**
Image #9	0.90	0.88	0.85	0.96	**0.9906**
Image #10	0.91	0.90	0.89	0.92	**0.9906**
Image #11	0.92	0.89	0.91	**0.97**	**0.9738**

Table 6: Quantitative results on the whole database

Image #	ER	GCE	PRI	VI
ISVM	6.13	0.14	0.91	0.26
FSVM	7.25	0.17	0.88	0.25
FCM-SVM	8.11	0.22	0.87	0.31
SSVM	4.26	0.07	**0.92**	**0.21**
The proposed method	**3.91**	**0.05**	**0.92**	0.27

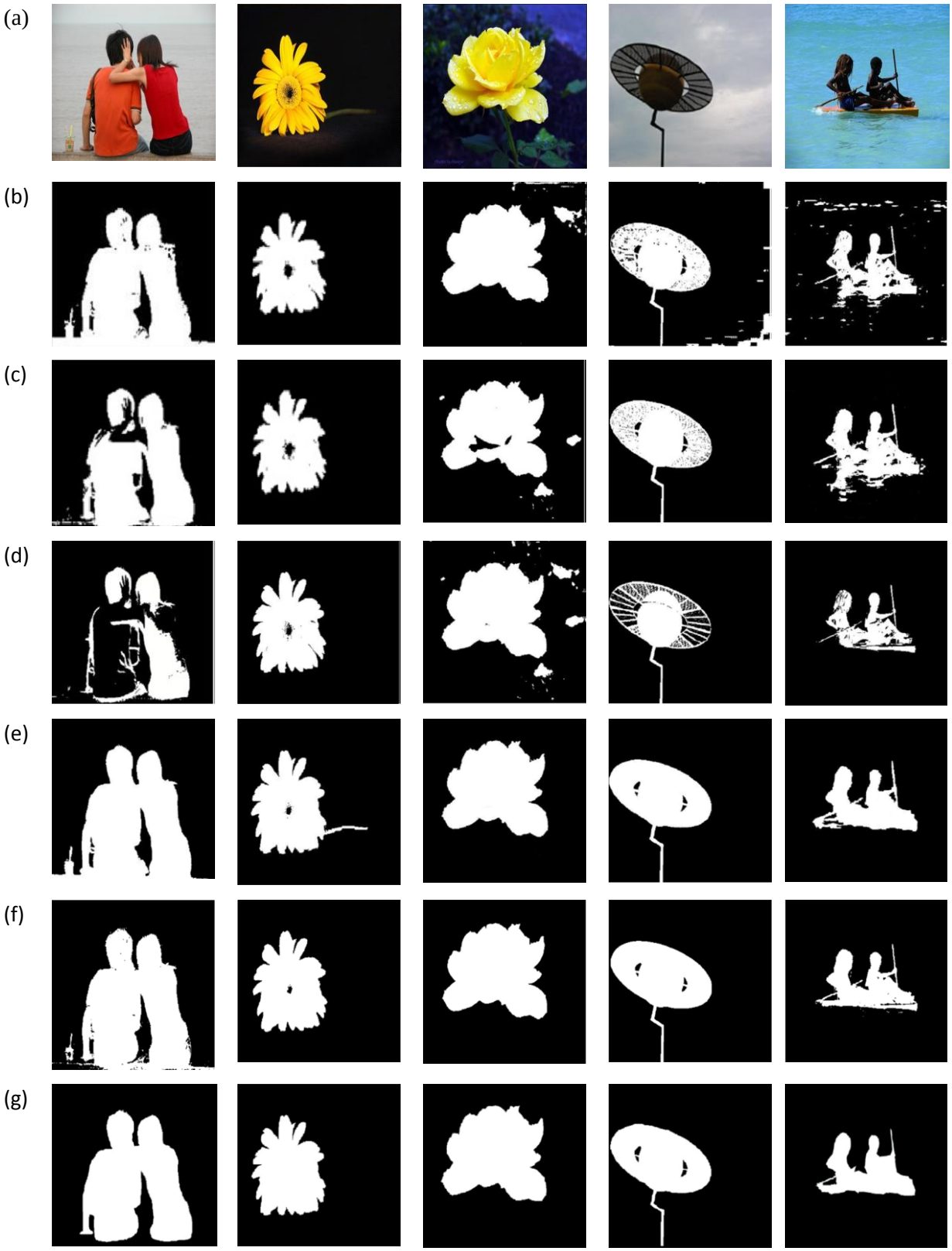


Fig. 3. (a) original images; (b) segmentation results of ISVM (c) FSVM; (d) FCM-SVM; (e) SSVM; (e) the proposed method; (g) ground truth segmentations

References

[1] X. Y. Wang, W. W. Sun, Z. F. Wu, H. Y. Yang, "Color image segmentation using PDTDFB domain hidden Markov tree model", Applied Soft Computing, Vol. 29, 2015, pp. 138-152.

[2] A. Dirami, K. Hammouche, M. Diaf, P. Siarry, P., "Fast multilevel thresholding for image segmentation through a multiphase level set method", Signal processing, 93(1), 2013, pp. 139-153.

[3] H. Cai, Z. Yang, X. Cao, W. Xia, X. Xu, "A new iterative triclass thresholding technique in image segmentation", IEEE transactions on image processing, Vol. 23, No. 3, 2014, pp.1038-1046.

[4] L. U. Ambata and E. P. Dadios, "Foreground Background Separation and Tracking", International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment (HNICEM), 2019, pp. 1-6.

[5] F. A. Khan, M. Nawaz, M. Imran, A. U. Rahman and F. Qayum, "Foreground detection using motion histogram threshold algorithm in high-resolution large datasets", Multimedia Systems, 2020, pp. 1-12.

[6] M. Castillo-Martinez, F. J. GallegosFunes, B. E. Carvajal-Gamez, G. Urriolagoitia-Sosa, A. J. Rosales-Silva, "Color index based thresholding method for background and foreground segmentation of plant images", Computers and Electronics in Agriculture, Vol. 178, 2020, p. 105783.

[7] J. Canny, "A computational approach to edge detection", IEEE Transactions on pattern analysis and machine intelligence, Vol. 93, No. 6, 1986, pp. 679-698.

[8] J. M. Prewitt, "Object enhancement and extraction", Picture processing and Psychopictorics, Vol. 10, No. 1, 1970, pp. 15-19.

[9] R. C. Gonzalez and R. E. Woods, "Digital image processing", 2002.

[10] T. Uemura and G. Koutaki and K. Uchimura, "T. Uemura and G. Koutaki and K. Uchimura", International Journal of Innovative computing, Information and control, Vol. 7, No. 10, 2011, pp. 6073-6083.

[11] D. Díaz-Pernil, A. Berciano, F. Peña-Cantillana and M. A. Gutiérrez-Naranjo, "Segmenting images with gradient-based edge detection using membrane computing", Pattern Recognition Letters, Vol. 34, No. 8, 2013, pp. 846-855.

[12] C. Panagiotakis, I. Grinias and G. Tziritas, "Natural image segmentation based on tree equipartition, bayesian flooding and region merging", IEEE Transactions on Image Processing, Vol. 20, No. 8, 2011, pp. 2276-2287.

[13] J. Ning, D. Zhang and C. Wu and F. Yue, "Automatic tongue image segmentation based on gradient vector flow and region merging", Neural Computing and Applications, Vol. 21, No. 8, 2012, pp. 1819-1826.

[14] L. Wang, H. Wu and C. Pan, "Region-based image segmentation with local signed difference energy", Pattern Recognition Letters, Vol. 34, No. 6, 2013, pp. 637-645.

[15] S. E. Ebadi and E. Izquierdo, "Foreground segmentation with tree-structured sparse RPCA", IEEE transactions on pattern analysis and machine intelligence, Vol. 40, No. 9, 2017, pp. 2273-2280.

[16] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts", IEEE Transactions on pattern analysis and machine intelligence, Vol. 23, No. 11, 2001, pp. 1222-1239.

[17] T.M. Nguyen and Q. J. Wu, "Fast and robust spatially constrained Gaussian mixture model for image segmentation", IEEE transactions on circuits and systems for video technology, Vol. 23, No. 4, 2012, pp. 621-635.

[18] O. O. Karadag and F. T. Y. Vural, "Image segmentation by fusion of low level and domain specific information via Markov Random Fields", Pattern Recognition Letters, Vol. 46, 2014, pp. 75-82.

[19] N. Dhanachandra, K. Manglem and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm", Procedia Computer Science, Vol. 54, 2015, pp. 764-771.

[20] Y. Yang, Y. Wang and X. Xue, "A novel spectral clustering method with superpixels for image segmentation", Optik, Vol. 127, No. 1, 2016, pp. 161-167.

[21] L. A. Lim and H. Y. Keles, "Foreground segmentation using convolutional neural networks for multiscale feature encoding", Pattern Recognition Letters, Vol. 112, 2018, pp. 256-262.

[22] A. Shahbaz and K. H. Jo, "Deep Foreground Segmentation using Convolutional Neural Network", IEEE 28th International Symposium on Industrial Electronics (ISIE), 2019, p. 103334.

[23] P. Patil and S. Murala, "Fggan: A cascaded unpaired learning for background estimation and foreground segmentation", IEEE Winter Conference on Applications of Computer Vision (WACV), 2019, pp. 1770-1778.

[24] D. Sakkos and E. S. Ho and H. P. Shum, "Illumination-aware multi-task GANs for foreground segmentation", IEEE Access, Vol. 7, 2019, pp. 10976-10986.

[25] J. Liang and Y. Xue and J. Wang, "Genetic programming based feature construction methods for foreground object segmentation", Engineering Applications of Artificial Intelligence", Vol. 89, 2020, p. 103334.

[26] Z. Yu, H. S. Wong and G. Wen, "A modified support vector machine and its application to image segmentation", Image and Vision Computing, Vol. 29, No. 1, 2011, pp. 29-40.

[27] X. Y. Wang, Q. Y. Wang, H. Y. Yang and J. Bu, "Color image segmentation using automatic pixel classification with support vector machine", Neurocomputing, Vol. 74, No. 18, 2011, pp. 3898-3911.

[28] X. Bai and W. Wang, "Saliency-SVM: An automatic approach for image segmentation", Neurocomputing, Vol. 136, 2014, pp. 243-255.

[29] M. K. Sangale and N. B. Kadu, "Real-time Foreground Segmentation and Boundary Matting for Live Videos using SVM".

[30] C. Tang and M. O. Ahmad and C. Wang, "Foreground segmentation in video sequences with a dynamic background", 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2018, pp. 1-6.

[31] L. U. Shuhan and S. J. YE, "Using an image segmentation and support vector machine method for identifying two locust species and instars", Journal of Integrative Agriculture, Vol. 19, No. 5, 2020, pp. 1301-1313.

[32] N. Dhanachandra, K. Manglem, and Y. Chanu, "Image segmentation using K-means clustering algorithm and

subtractive clustering algorithm". Procedia Computer Science, 54, 2015, pp. 764-771.

[33] W. Chen, C. He, C. Ji, M. Zhang, S. and Chen, "An improved K-means algorithm for underwater image background segmentation", Multimedia Tools and Applications, 80(14), 2021, pp. 21059-21083.

[34] Y. Yang, H. Bilen, Q. Zou, W. Y. Cheung, X. Ji, "Learning Foreground-Background Segmentation from Improved Layered GANs", In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2524-2533), 2022.

[35] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers", In Proceedings of the fifth annual workshop on Computational learning theory, 1992, pp. 144-152.

[36] V. N. Vapnik, "Statistical learning theory", Wiley, New York, 1998.

[37] M. A. Aizerman,"Theoretical foundations of the potential function method in pattern recognition learning",Automation and remote control, Vol. 25, 1964, pp.821-837.

[38] S. Lloyd,"Least squares quantization in PCM", IEEE transactions on information theory, Vol. 28, No. 2, 1982, pp. 129-137.

[39] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression", IEEE transactions on image processing, Vol. 19, No. 1, 2009, pp. 185-198.

[40] P. Soille, "Morphological image analysis: principles and applications", Springer Science and Business Media, 2013.

[41] R. Achanta and Sh. Hemami and F. Estrada and S. Susstrunk, "Frequency-tuned salient region detection", IEEE conference on computer vision and pattern recognition, 2009, pp. 1597-1604.

[42] R. Unnikrishnan, C. M. Pantofaru and M. Hebert, "Toward objective evaluation of image segmentation algorithms", IEEE transactions on pattern analysis and machine intelligence, Vol. 29, No. 6, 2007, pp.929-944.

[43] M. Meila,"Comparing clusterings—an information based distance",Journal of multivariate analysis, Vol. 98, No. 5, 2007, pp. 873-895.

[44] D. Martin and C. Fowlkes and D. Tal and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vol. 2, 2001, pp. 416-423.