ISSN 2210-142X

International Journal of Computing and Digital Systems (IJCDS)

VOLUME NO. 14
ISSUE NO. 2
MAY- AUGUST 2025

JE - 18,Gupta Colony, Khirki Extn,
Malviya Nagar, New Delhi - 110017.
E- Mail: info@enrichedpublication.com
Phone :- +91-8877340707

ISSN 2210-142X

International Journal of Computing and Digital Systems (IJCDS)

Aims and Scope

International Journal of Computing and Digital Systems (IJCDS) is a peer-reviewed International Journal that currently publishes 6 issues annually. IJCDS journal publishes technical papers, as well as review articles and surveys, describing recent research and development work that covers all areas of computer science, information systems, and computer / electrical engineering.

The topics covered by IJCDS are including and not limited to the following research areas:

- Reconfigurable Computing & Embedded systems
- Computer Communications and Networking
- Internet of Things & Real Time Systems
- Cyber Security
- Cloud Computing
- Smart Systems
- Information Systems and Communication Service
- Innovation/Technology Management
- Business Information Systems
- Software Engineering
- Mobile & Web Applications
- Theory of Computation
- Data Structures, Cryptology and Information Theory
- Artificial Intelligence & Robotics
- Image Processing, Computer Vision, Pattern Recognition & Graphics
- Data Mining & Big Data
- Smart Grids & Renewable Energy
- Human Computer Interaction

Responsiveness Speed

Time from submission to first decision after peer review: 4-8 Weeks
Time to immediate reject: 2-4 Weeks

Publication Speed

Time from final acceptance to print: 3-6 Months
Total time to publication: 4-8 Months
Time from submission to acceptance: 4-12 Weeks
Time from acceptance to online Publishing: 4-8 Weeks

Publishing Ethics

Plagiarism in all its forms constitutes unethical publishing behavior and is unacceptable. The journal has no tolerance on plagiarism. All submitted manuscripts must go through cross checking using turnitin as an online plagiarism checker.

ISSN 2210-142X

International Journal of Computing and Digital Systems (IJCDS)

(Volume No. 14, Issue No. 2, May- August 2025)

Contents

Predicting the Import and Export of Commodities using Support Vector Regression and Long Short-Term Prediction Models

Razia Sulthana A[1], Pranav Ramesh[2]
[1] Department Of Computer Science And Engineering, BITS Pilani, Dubai, United Arab Emirates
[2]Department Of Computer Science And Engineering, BITS Pilani, Dubai, United Arab Emirates
E-mail address: razia@dubai.bits-pilani.ac.in, f20180089@dubai.bits-pilani.ac.in

ABSTRACT

The prediction of import and export of commodities have occurred between countries to either buy or sell goods essential for humans. Governments need to keep track of the amount of import or export to ensure the increase of Gross Domestic Product (GDP) for their country. Support Vector Machine (SVM) is a powerful classification algorithm to classify data efficiently. Support Vector Regression (SVR) is a modification of SVM that predicts absolute values. The purpose of this paper is to use SVR in a commodity dataset to predict each commodity's price being imported and exported for limited countries. SVR uses the support vectors obtained during the running of the algorithm to predict the dataset's outcome. The new version of SVR algorithm is proposed which is assisted with modified RBF Kernel to improve the model's efficiency. Further LSTM is applied for prediction in layers to predict the weight of some incoming commodities to countries. We then obtain the predicted results and find the accuracy of the model using this result over a real dataset. The results show that the over-all error for the proposed model is very trivial and hence produces higher accuracy.

Keywords: Support Vector Machine, Regression model, Commodity Prediction, Kernels, LSTM.

1. INTRODUCTION

Analyzing the import and export of a country is crucial for the Gross Domestic Product (GDP) of the country [1]. A country's GDP decides the economic stability and values the country amidst others. Excessive import to a country harms the GDP for the importing country. On the contrary, excessive export from a country has a positive impact on the exporting country's GDP. It is important to keep track of commodities entering and exiting to ensure the merchandise received or given is safe and to ensure the product received can be taken by all goods transporters available on that day [2].

By predicting the amount of import and export, it helps identify how many carriages should be there at each port and uses methods to reduce the amount of time for planning, the cost for traveling, carriage etc. The study of influential factors to GDP and the patterns in then would help in keeping up a country stable, which however requires intelligent prediction algorithms.

Machine Learning operates on artificial learning techniques that trains a machine to learn on its own from the previously acquired knowledge [3]. Artificial Intelligence (AI) is used in solving complex tasks to ease labor. Machine Learning has many algorithms for forecasting, image processing, facial and sound recognition, natural language processing (NLP), etc. This paper will focus on applying on a supervised learning algorithm, Support Vector Machine (SVM) (Fig. 1).

SVM, a powerful classification algorithm developed from statistical learning. It was proposed by Vapnik in 1963 and has attracted many researchers to apply this model to various datasets ever since. This algorithm is subjected to fluctuation in performance based on how the cost parameter and the programmers set the kernel parameters [4]. Hence, in this paper an extensive cross-validation is performed to find an optimal parameter setting.

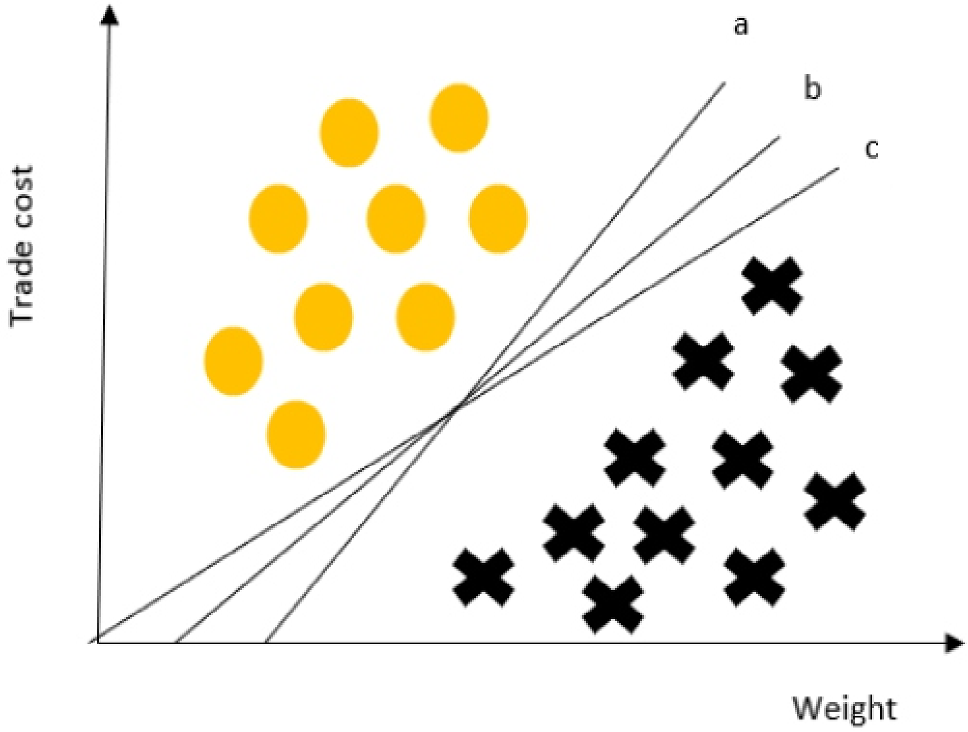


Figure 1. Hyperplanes generated by SVM for classifying data points

Vapnik built another form of the supervised algorithm called Support Vector Regression (SVR). This algorithm proves to be a very efficient algorithm compared to other regression algorithms due to a kernel that instructs the algorithm to predict the data in any fashion. The high volume of trade along with its multidimensional features can be supervised by SVR. In general, Import means to shop for goods and services from a distinct country to the house country. Therefore, these goods and services are produced in an exceedingly foreign land and bought by the actual domestic country. Export means transporting goods or services to a different country. As a result, it brings within the foreign income to the domestic country. The sale of goods adds to the nation's gross output. On applying SVR the future import and export values of certain commodities can be appropriately predicted.

Long Short-Term Memory (LSTM) is another deep learning neural network. It can process multiple unstructured data from the input gate. It is used to infer information based on data that has a time series. This was initially developed to remove the problems that occurred due to gradient descent. Over time this is used for image, speech, and handwriting recognition and is very powerful when analyzing stock market data. This proves to be a powerful recurrent neural network (RNN) compared to other RNN. It is well known to make accurate predictions on time-series datasets such as stock market, temperature predictions, etc.

This paper focuses on applying a supervised learning algorithm, SVR, on a commodity dataset to predict the output for different inputs parameters. The proposed approach finds the optimized value of the goods' cost and the weight of the goods entering and leaving to ensure that commodities are traded safely. Following which it applies LSTM to predict the weight of some commodities coming into countries.

The section split of the paper is given here. Section 2 discuss in detail the literature review related to the proposed work.

2. LITERATURE REVIEW

In this section, we will see some major augmentations towards SVM, SVR, and LSTM. Yingjie et al. [5] in their paper has shown that SVM can be used not only in solving simple optimization problems, for example, Linear Programming, Quadratic Programming, etc., but rather can solve more general optimization problems, for example, integer programming, semi-infinite programming, etc. They have also proposed a real-time application of these general optimization problems in the field of economics. The author in [6] have shown that the RBF the kernel has shown maximum accuracy Linear, Polynomial, and Sigmoidal on various datasets collected by them, for example, MIT, Yale, etc. Another article [7] have shown SVM's application in Data Mining and how essential it is to this field by providing various literature reviews. Also, they have provided an application for various SVM models.

Various known kernel methods have been compared in [8] and have shown how each kernel could be used in termite detection. They found that the polynomial kernel had better accuracy than the other kernel methods for termite detection.

Vikas et al. [9] have compared various kernel on remotely sensed satellite data gathered from QuickBird and Landsat Enhanced Thematic Mapper Plus (ETM+) and have shown that for QuickBird polynomial kernel performed the best and for the ETM+ sigmoid kernel gave the best accuracy. However, they also proposed larger data. It depends on how well we optimize the parameters for the kernel functions to get the best accuracy.

In [10] kernel performance on a multi-class vowel data is compared and has shown that the RBF with 36-dimensional MFCC data performs better than the remaining kernels. Laura and Rouslan [11] have used SVM in Solvency Analysis and compared this to traditional approaches, such as logistic regression and discriminant analysis. They have shown that SVM can be an alternative to measure the company rating. A new and better kernel function is proposed in [12] called Radial Based Polynomial Kernel (RBPK). This kernel combines the characteristics of both the Polynomial and Radial Based Function. They have proved the new kernel to be a valid kernel and a more accurate kernel than Linear, Polynomial, and Radial Based Function Kernels individually.

In a similar way, support vector regression along with RBF kernel and multiple linear regression predicts the absorption rate of lead (II) ions. They show that Support Vector Regression provides a near perfect result compared to multiple linear regression [13]. SVR is also applied in [14] and have given an overview of SVR and different areas from where this concept is derived from. They also give new ideas which have emerged from this Support Vector Regression which is still in research in current times.

Support Vector Regression is used for Newspaper and Magazine Sales and compares this model using linear and RBF kernel [15]. Both showed equally good results and are dominant when predicting the sales of newspaper and magazine. An enhanced support vector regression is developed in [16] with more un-interpretable kernel and has used this model for weather forecasting

A number of research work is done in stock prediction using SVM and SVR. The in [17] predict the stock prices for IBM INC. from historical data using an old machine algorithm known as Support Vector Machine. The model is aided with the kernel, radial basis function, to gain more data accuracy. A systematic study of various journal papers is done in [18] to comprehend the machine learning algorithm that will be most effective for stock market predictions.

In [19], the authors provide a deep learning model to predict Chinese stock market data's stock prices. The authors developed a new algorithm known as feature extension. Finally, they compared their deep learning model to traditional machine learning models. The results clearly showed that the deep learning model was advantageous as compared to conventional ML models, and it is a topic for further study.

A Random forest technique to predict stock prices with actual data and sentiment data is proposed in [20] which proves the effectiveness of random forest on various volatile data. Hence its erratic to say that a specific ML and DL algorithms will not suit for all problems.

In [21], the authors use and compare the prediction made by Linear Regression, Exponential Smoothing, and Time Series Forecasting on Amazon, AAPL, and Google stocks obtained from Yahoo Finance. The results show that the exponential smoothing provides lesser error and higher accuracy compared to the others. The authors in [22] compare the performance of Linear Regression, Polynomial Regression, and Support Vector Regression on stock prices of the S&P 500. The results showed that Support Vector Regression outperformed as compared to Linear and Polynomial regression.

LSTM is applied in [23] to foresee the future pattern that might take place on stock prices based upon a dataset. The results obtained had 55.9% accuracy when checked where the stock will go in the future, up or down. Also, a model that predicts how long will stock will last and how essential it will be for traders is given in [24]. The accuracy of the random forest shows higher value than that of the Support Vector Classifier. Artificial neural networks and support vector machines (SVM) and applied together to predict future stock market prices [25]. A comparative work over the best ML algorithm to predict the stock market data is shown in [26]. Once again, random forest produced the highest accuracy and recall rate as compare to KNN.

3. METHODOLOGY

The implementation model is done by analyzing the import and export data separately. As a first step of implementation, the dataset is split into import and export data by splitting the training and test data. This is to ensure to avoid overfitting.

Step 1: Search for data records that has flow which is 'import' and flow that is 'export' and assign it to separate variables as shown below.

Step 2: To split the dataset a function present python is applied to implicitly split the data randomly into train and test data.

A. SVM and SVR architecture diagram

The dataset contains 8225871 rows × 10 columns taken from Kaggle and have received the data from the United Nations Statistics Division website. The link to the dataset: https://www.kaggle.com/united nations/global-commodity-trade-statistics#commodity_trade_statistics_data.csv. The dataset is about the import and export of each country for different commodities and the trade price of each commodity (Fig. 2).

```
[ ]  data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8225871 entries, 0 to 8225870
Data columns (total 10 columns):
 #   Column           Dtype
---  ------           -----
 0   country_or_area  object
 1   year             int64
 2   comm_code        object
 3   commodity        object
 4   flow             object
 5   trade_usd        int64
 6   weight_kg        float64
 7   quantity_name    object
 8   quantity         float64
 9   category         object
dtypes: float64(2), int64(2), object(6)
memory usage: 627.6+ MB
```

Figure 2. Attribute information of the table

Support Vector Machines is used to classify the data into two or more classes based on their similarities to that class, with known class labels. This algorithm classifies the data by drawing a hyperplane that splits the datasets and ensure each class is as far as possible from the decision boundary. Given below is the architecture diagram shown in the Fig. 3.

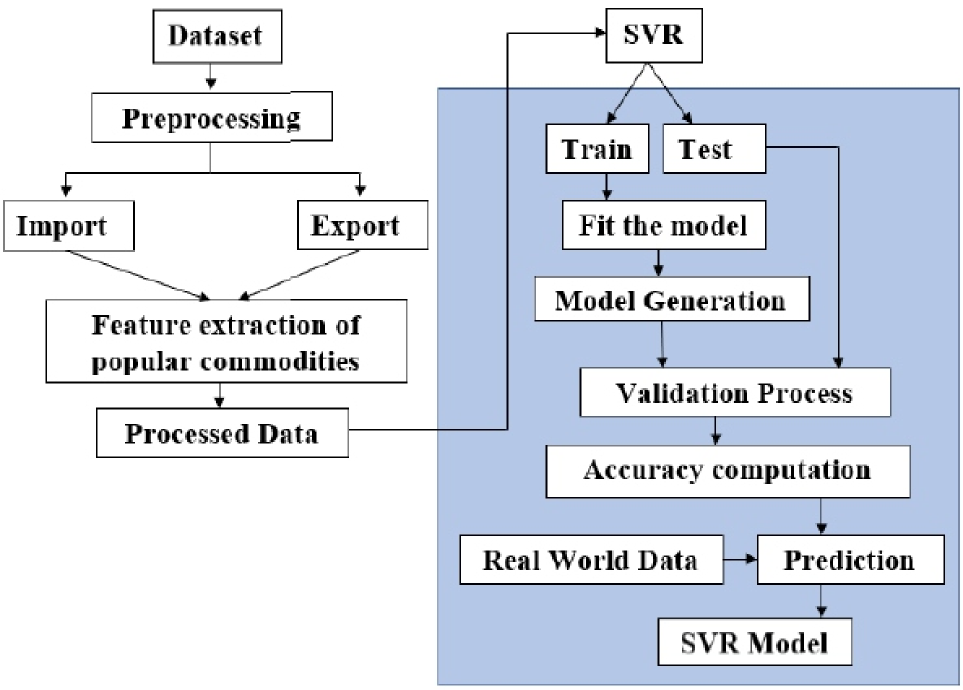


Figure 3. Architecture diagram of SVR Model

Step 1: Project $\vec{x}$ onto $\vec{w}$ and set is greater than or equal to a constant c

$$\vec{x}.\vec{w} \geq c \qquad (1)$$

Step 2: Without losing generality (WOLOG)

$$\vec{x}.\vec{w} - c \geq 0 \qquad (2)$$

$$\vec{x}.\vec{w} + w_0 \geq 0, where\ w_0 = -c \qquad (3)$$

Step 3: The above equations for a discriminant function, now for SVM.

$$\overrightarrow{x_+}.\vec{w} + w_0 \geq 1, \text{if } y_i = 1 - Eq\ 1 \qquad (4)$$

$$\overrightarrow{x_-}.\vec{w} + w_0 \leq -1, \text{if } y_i = -1 - Eq\ 2 \qquad (5)$$

Step 4: Now we have two equations, due to the fact that there are two classes present in different areas of the data, but it is better to have one equation that can work for both cases.

Multiply (4) with y_i in the LHS side

$$y_i\ (\overrightarrow{x_+}.\vec{w} + w_0) \geq 1 - Eq\ 3 \qquad (6)$$

Multiply (5) with $-\ y_i$ in the LHS side

$$-y_i\ (\overrightarrow{x_-}.\vec{w} + w_0) \leq -1 \qquad (7)$$

Rearranging we get,

$$y_i\ (\overrightarrow{x_-}.\vec{w} + w_0) \geq 1 - Eq\ 4 \qquad (8)$$

Now we see that (6) and (8) are the same only difference is the point which can be generalized.

$$y_i\ (\vec{x}.\vec{w} + w_0) \geq 1, where\ i = 1, 2, \ldots, N \qquad (9)$$

$$y_i\ (w^T\mathrm{x} + w_0) \geq 1, where\ i = 1, 2, \ldots, N \qquad (10)$$

This is due to the fact in Machine learning the algorithm performs the above step more frequently on the data. Now, x can be processed to something more useful and we call that $\phi(x)$

Where, $\phi: R^n \rightarrow R^n$ is a mapping from features to is a mapping from features to high dimensional feature space, where these points can become linearly separable.

The distance of the point xi the hyperplane is given by:

$$d(x_i) = \frac{y_i(x)}{||w||^2} \qquad (11)$$

$$d(x_i) = \frac{|w^T \phi(x_i) + w_0|}{||w||^2} \qquad (12)$$

To find the optimal hyperplane that separates the data is by solving the following optimization problem:

$$\min \varphi(w) = \frac{1}{2}||w||^2 \qquad (13)$$

The solution to the above problem is given by the saddle point of the Lagrange formula

$$\mathrm{L}_{P1} = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i [y_i (w^T \phi(x_i) + w_0) - 1], \qquad \alpha_i \geq 0 \qquad (14)$$

So far, we have based on the assumption that the data is separable. Now, to apply this to a general case we introduce a slack variable ξi such that

$$y_i (w^T \phi(x_i) + w_0) \geq 1 - \xi_i, where\ \xi_i \geq 0\ and\ i = 1, 2, \dots, N \qquad (15)$$

We add an additional cost to the objective function as there is an upper bound to the slack variable.

$$\min \varphi(w, \xi) == \frac{1}{2}||w||^2 + C \sum_{i=1}^{N} \xi_i \qquad (16)$$

$$\mathrm{L}_{P2} = \frac{1}{2}||w||^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i [y_i (w^T \phi(x_i) + w_0) - 1 + \xi_i] - \sum_{i=1}^{N} \beta_i \xi_i \qquad (17)$$

$$where\ \alpha_i, \beta_i \geq 0$$

Applying the Karush–Kuhn–Tucker (KKT) conditions we get

$$\frac{\partial \mathrm{L}_{P2}}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y_i\, \phi(x_i) = 0 \qquad (18)$$

$$\frac{\partial \mathrm{L}_{P2}}{\partial w_0} = -\sum_{i=1}^{N} \alpha_i y_i \ = 0 \tag{19}$$

$$\frac{\partial \mathrm{L}_{P2}}{\partial w_0} = \ C - \ \alpha_i - \ \xi_i \tag{20}$$

$$y_i \ (w^T \phi(x_i) + w_0) \ \geq 1 - \ \xi_i \tag{21}$$

$$\xi_i, \alpha_i, \beta \geq 0$$

$$\alpha_i [y_i \ (w^T \phi(x_i) + w_0) - 1 + \ \xi_i \ ] = 0 \tag{22}$$

$$\beta_i \xi_i \ = 0$$

Hence, we get

$$w = \sum_{i=1}^{N} \alpha_i y_i \ \phi(x_i) \tag{23}$$

We get an interesting observation that ξ_i=0 to calculate w_0

$$w_0 = y_j - w^T \phi\left(x_j\right) \tag{24}$$

For numerical reasons we use mean and get

$$w_0 = \frac{1}{N_s} \sum_{0<\alpha_i<C} y_j - w^T \phi\left(x_j\right), N_s \ is \ no \ of \ support \ vectors \tag{25}$$

For a new data we use the following function

$$f(x) = Sign(w^T \phi(x) + w_0) \tag{26}$$

$$f(x) = Sign\left(w^T \phi(x) + \frac{1}{N_s} \sum_{0<\alpha_i<C} y_j - w^T \phi\left(x_j\right) \right) \tag{27}$$

$$(x) = Sign\left(\sum_{i=1}^{N} \alpha_i y_i \, \phi(x_i)^T \, \phi(x) + \frac{1}{N_s} \sum_{0<\alpha_i<C} y_j - \alpha_i y_i \, \phi(x_i)^T \phi(x_j)\right) \quad (28)$$

And if we use a kernel then

$$k(x_i, x) = \phi(x_i)^T \phi(x) \quad (29)$$

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (30)$$

Kernels in SVM helps to convert the data to a high dimensional space and thereby ensures classifying the data using linear or non-linear boundaries.

Kernels: A few of the regular used kernel types and equations are listed here.

1. Polynomial, $k(x_i, x_j) = (x_j^T x_i + t)^d$
2. Gaussian Radial Basis Kernel, $k(x_i, x_j) = e^{\frac{-||x_i - x_j||^2}{\sigma^2}}$
3. Sigmoid Kernel, $k(x_i, x_j) = \tanh(x_j^T x_i + t)$

C. Long Short-Term Memory

The input passes to the input gate of the LSTM model (Fig. 4). It passes through various layers in the model and passing through layers comes with a model with a reasonable prediction rate. This model is used to predict the test data set, and we get the model's accuracy with the help of the R2 score.

The equations at the gate are

Input Gate: $i_g = sigmoid(w_i . input + b_i)$

Output Gate: $o_g = sigmoid(w_o . input_o + b_o)$

Final Output: $h_t = o_g * \tanh(c^t)$, c^t represents cell state at time t.

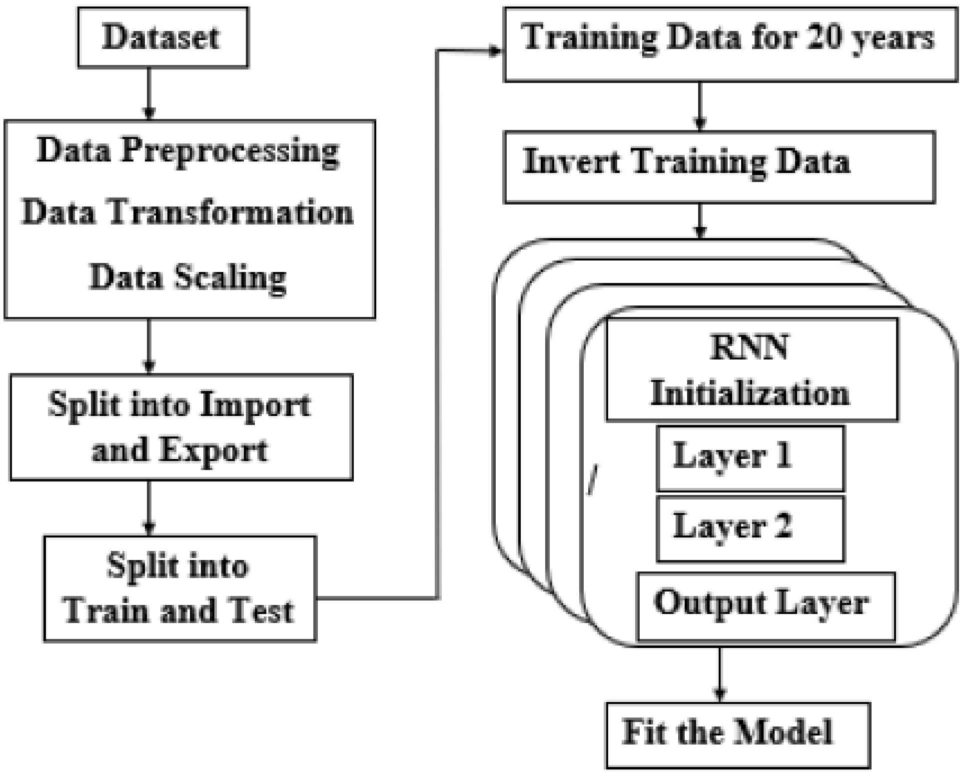


Figure 4. LSTM architecture

D. Performance Measure

1) R2 score: Here, we will use square of Pearson product-moment correlation coefficient (R2) score as the measure of performance of the model. This is an important measure of degree for regression problems. Many researchers refer this as correlation coefficient. R2 is defined as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad (31)$$

Where, SSR is sum of squared regression
SST is sum of squared total
SSE is sum of squared error

It measures the amount of variance in dependent variable compared from the independent variable. Python provides a beautiful function present in sklearn. metrics module called r2_score which implicitly does this calculation, also the score() function of the model by default uses R2 score.

R2 score is generally in the range between 0 and 1, where a value of 1 indicates the prediction of the regression model is perfect. However, if it is out of this range, then the model is worse than a horizontal hyperplane.

2) Root Mean Square Error (RMSE): RMSE is a metric used to measure the performance of various models. It is derived by taking the square root of the mean squared error. Mean squared error is obtained by squaring the difference between the predicted value from the target value and dividing the result by the total number of training dataset. This is a distinctive measure that tells the machine algorithm to reduce the difference between predicted and target values for the values to come closer to each other. It is given by the formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{31}$$

N – No. of training examples

y_i – Target Value

$\hat{y}_i$ – Predicted Value

4. RESULTS

The models are analyzed with different countries for various commodities. As a first step export data is analyzed and following which import data is analyzed.

A. Export - Aircraft parts necessary

This commodity represents the amount of aircraft parts required by the country to make airplanes. Here is the table for Aircraft parts commodities only

Now we are going to analyze for the following countries (Table 1):

1. United Arab Emirates
2. Australia
3. Japan
4. India

The analysis is between weight of the commodity vs trade cost of the commodity using SVR and RBF Kernel. The following Table 1 shows the parameters for each country taken by the model to try and perfectly fit the model along with the R2 score for each of them.

TABLE 1. R2 value of export values of different countries – Aircraft Parts

Country	c	Gamma	Epsilon	R^2 Score (For Test)	R^2 Score (For Train)
UAE	10E10	1.00E-10	0.1	0.871285262	0.82032062
Switzerland	4.1979 *10E8	10E-B	0.1	0.583154081	0.925789726
Japan	9.1029 *10E11	4.0949x 10E-15	0.1	0.972947001	0.983540155
India	3.089 *10E9	5.6898 *10E-13	0.1	0.868845563	0.985850276

From the above table we make the following observations:

1. For the country UAE the data nearly fits the model perfectly with R^2 score of 0.82032062 for the train and 0.871285262 for the test, with parameters C = 1010 and gamma = 10^{-11}. This shows that our model is perfect for the case of UAE (Fig. 5).
2. For the country Japan it gets a near perfect fit model with R^2score of 0.983540155 for the train and 0.972947001 for the test, with parameters C = 9.1029×10^{11} and gamma = 4.0949×10^{-15} . This shows that the model is perfect for Japan the only issue that can arise is overfitting (Fig. 5).

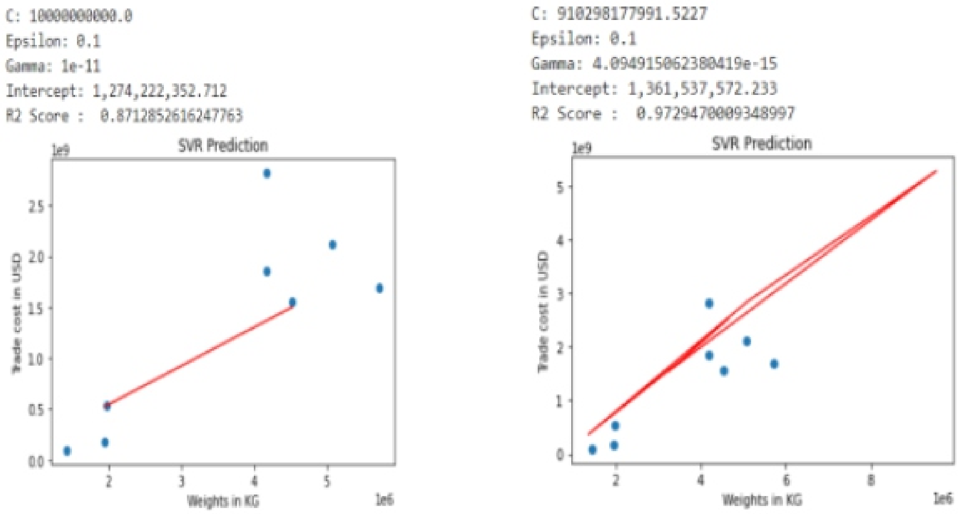


Figure 5. UAE and Japan export R^2values - Aircraft parts necessary

3. When it comes to the case of Switzerland the model over fits the data with R^2 score 0.583154081 for test and 0.925789726 for train (Fig. 6)
4. When we see India, the prediction is perfect with R^2 score 0.868845563 for test data and 0.985850276 for train. With parameters C = $3.089\ \ 10^9$ and gamma = $5.6898\ \ 10^{-13}$(Fig. 6)

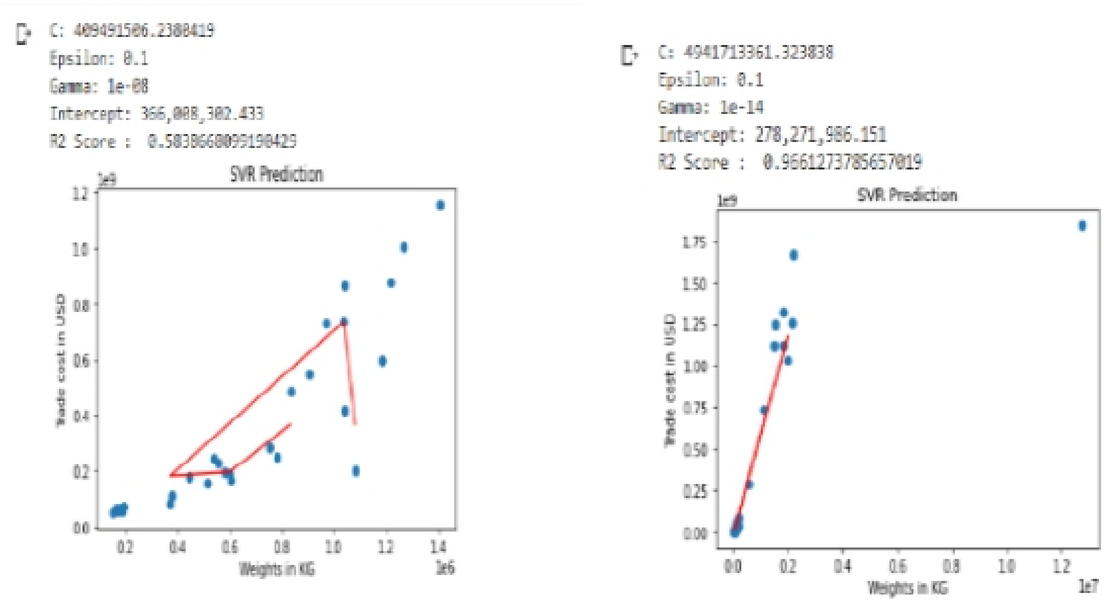


Figure 6. Switzerland and India export R^2values - Aircraft parts necessary

ISSN 2210-142X

B. Export - Food preparations necessary

The data over necessary food enquired for export is analyzed and R^2 values are obtained for countries (Table 2).

1. India
2. Republic of Korea
3. Japan
4. Portugal.

TABLE 2. R^2 value of export values of different countries – Food Preparation

Country	c	Gamma	Epsilon	R^2 Score (For Test)	R^2 Score (For Train)
India	1.0985* 10E9	1.00E-14	0.1	0.967032865	0.941698703
Rep. of Korea	6.2505* 10E9	1.00E-14	0.1	0.992613125	0.999309569
Japan	4.9417* 10E9	9.5409x l0E-14	0.1	0.836825022	0.987301212
Portugal	1.04811 *10E9	9.54095 x10E-15	0.1	0.830464177	0.882549269

For this commodity, the results were improbable.

1. For India, the prediction is perfect with R^2 score 0.967032865 for test and 0.941698703 for train.
2. For Korea it is a perfect fit model with R^2 score 0.992613125 for test and 0.999309569 for train (Fig. 7)

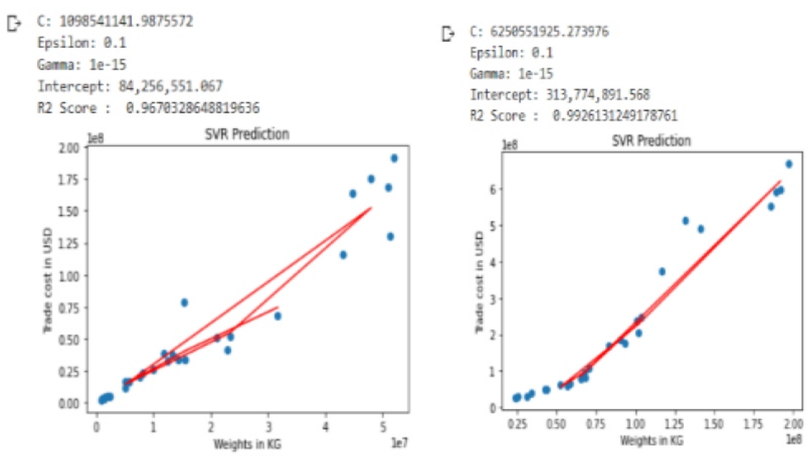


Figure 7. India and Korea export R^2 values - food preparations necessary

3. For Japan, R^2 score 0.836825022 for test and 0.987301212 for train.
4. For Portugal, a fine fit with R^2 score 0.830464177 for test and 0.882549269 for train (Fig. 8)

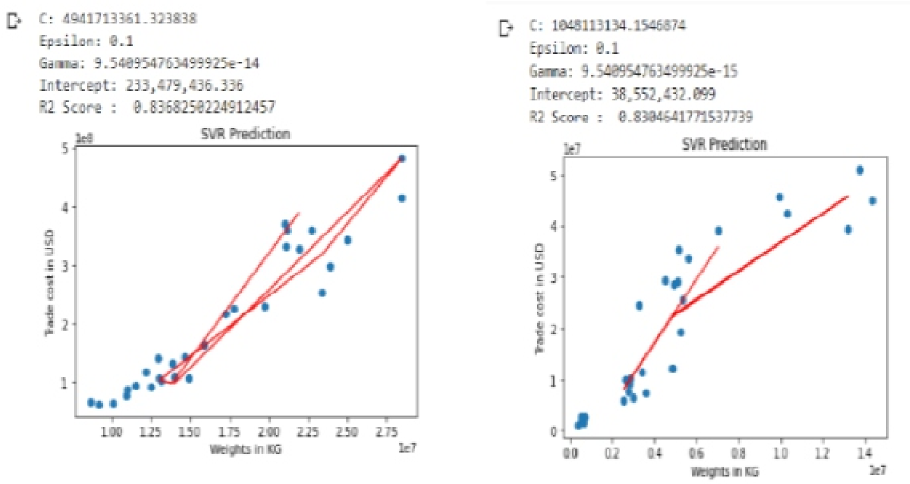


Figure 8. Japan and Portugal export R^2values -food preparations necessary

C. Import - Aircraft Parts Necessary

The import details of the aircraft parts required by the country to make airplanes is analyzed for UAE, Switzerland, Japan and Finland (Table 3)

TABLE 3. R^2 value of export values of different countries – Aircraft Parts Necessary

Country	c	Gamma	Epsilon	R^2 Score (For Test)	R^2 Score (For Train)
UAE	10E10	1.00E-14	0.1	0.48259007	0.649830112
Switzerland	8.90 x 10E9	1.00E-08	0.1	0.678877689	0.897543098
Japan	7.906xl0E11	1.00E-13	0.1	0.937966151	0.901391771
Finland	1.00E10	1.00E-11	0.1	0.692244907	0.812882187

1. For UAE R^2 score for test is 0.48259007 and train is 0.649830112 and Switzerland R^2score for test is 0.678877689 and train is 0.897543098 (Fig. 9).

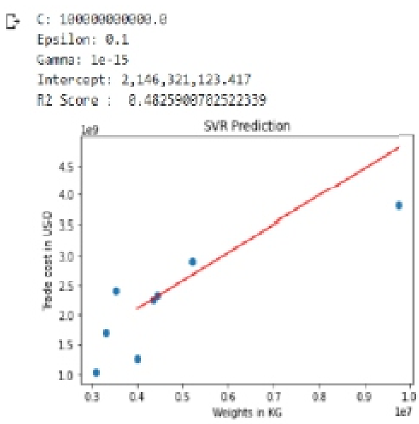


Figure 9. UAE import R^2 values – Air crafts necessary

2. For Japan R^2 score for test is 0.937966151 and train is 0.901391771 and for Finland the R^2 score for test is 0.692244907 and train is 0.812882187 (Fig. 10).

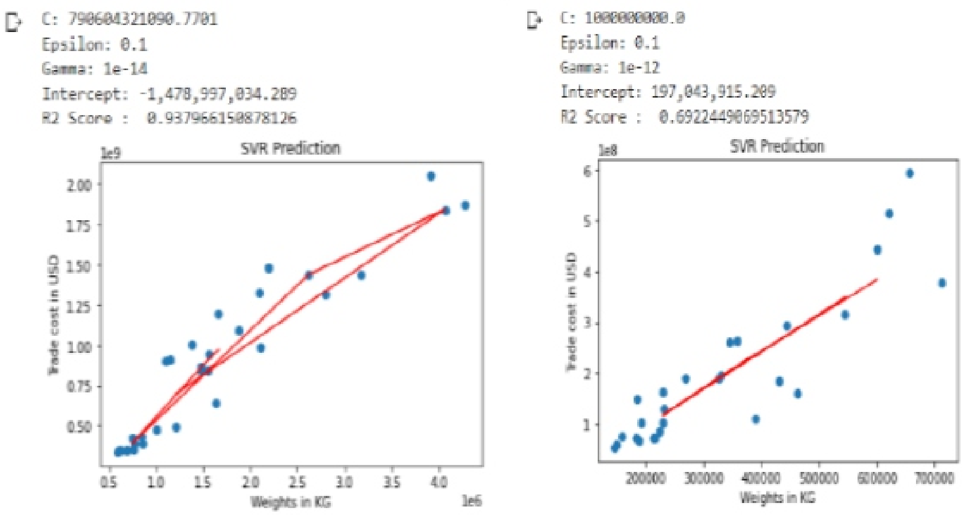


Figure 10. Japan and Finland import R^2values - Aircrafts necessary

D. Import -Food preparations necessary

The data over necessary food enquired for import is analyzed and R2 values are obtained for countries (Table 4).

1. India
2. Republic of Korea
3. Japan
4. Portugal.

TABLE 4. R^2 value of import values of different countries – Food Preparation

Country	c	Gamma	Epsilon	R^2 Score (For Test)	R^2 Score (For Train)
India	10E10	10E-15	0.1	0.418481896	0.456681645
Rep. of Korea	10E9	10E-14	0.1	0.971703336	0.973352841
Japan	10E9	10E-14	0.1	0.714932007	0.990669898
Portugal	1.9307 x10E9	10E-14	0.1	0.907346882	0.906586624

1. For India R2 score for test is 0.418481896 and train is 0.456681645.
2. For Rep. of Korea R2 score for test is 0.971703336 and train is 0.973352841 (Fig. 11).

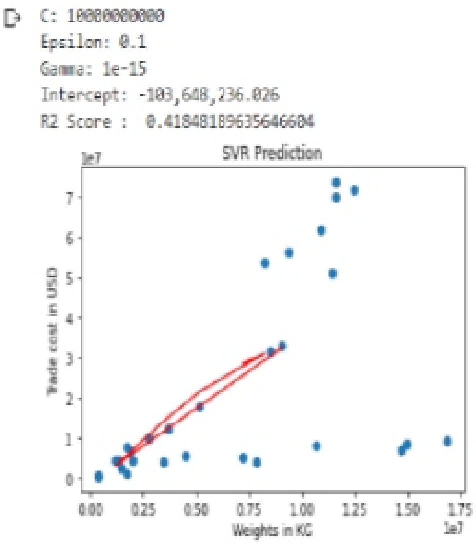


Figure 11. India and Republic of Korea import R^2values - Food preparations necessary

3. For Japan R^2 score for test is 0.714932007 and train is 0.990669898.
4. For Portugal R^2 score for test is 0.907346882 and train is 0.906586624 (Fig 12).

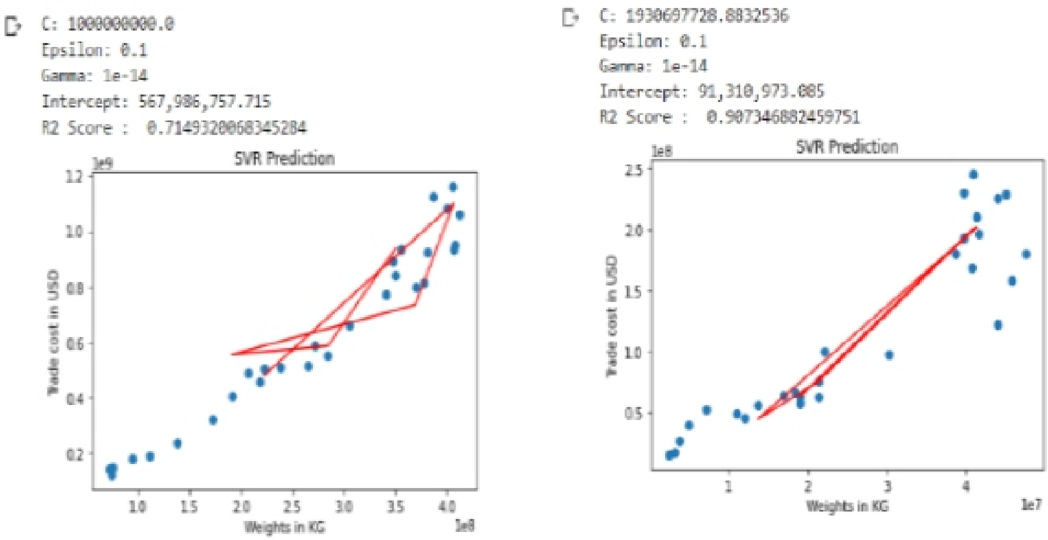


Figure 12. Japan and Portugal of Korea import R^2 values - Food preparations necessary

5. CONCLUSION

In this paper, the import and export values is analyzed using a joint approach using SVR and LSTM. The purpose of this paper is to use SVR in a commodity dataset to predict each commodity's price being imported and exported. SVR uses the support vectors obtained during the running of the algorithm to predict the dataset's outcome. The SVR algorithm is assisted with RBF kernel to improve the model's efficiency. The predicted results and the accuracy of the model obtained shows that the model produces better results with minimal error as compared to other models. The R^2 values tabulated signifies the model is good. This can be extended for predicting other logistics demand.

REFERENCES

[1] Dowrick, S., & Quiggin, J. (1997). True Measures Of Gdp And Convergence. The American Economic Review, 41-64.

[2] Shafaeddin, M. (1995). The Impact Of Trade Liberalization On Export And Gdp, Growth In Least Developed Countries. Unctad Review, 1995, 1-6.

[3] Mounicavennapusa And Srikanth Bhyrapuneni, "a Comprehensive Study Of Machine Learning Mechanisms On Big Data", International Journal Of Recent Technology And Engineering (ijrte), Madhya Pradesh, Issn: 2277-3878, Apr. 2 2019, Vol 7.

[4] D.shanthini, M.shanthi, Dr.m.c.bhuvaneswari, "a Comparative Study Of Svm Kernel Functions Based On Polynomial Coefficients And V-transform Coefficients", International Journal Of Engineering And Computer Science,madhya Pradesh, Issn:2319-7242, Mar 3 2017, Vol 6.

[5] Yingjie Tian, Yong Shi, Xiaohui Liu, "recent Advances On Support Vector Machines Research", Technological And Economic Development Of Economy, Issn: 2029-4913, Dec 19 2011, Vol 18.

[6] Deepika Kancherla, Jyostna Devi Bodapati, Veeranjaneyulu N, "effect Of Different Kernels On The Performance Of An Svm Based Classification", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Feb 2019, vol 7.

[7] Janmenjoy Nayak, Bighnaraj Naikand H. S. Behera, "A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges", International Journal of Database Theory and Application, Aug 1 2018, vol 8.

[8] Muhammad Achirul Nanda, KudangBoro Seminar, Dodi Nandika and AkhiruddinMaddu,"A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection", MDPI, Jan 2 2018.

[9] Vikas Sharma, Diganta Baruah, DibyajyotiChutia, PLN Raju, DK Bhattacharya, "An Assessment of Support Vector Machine Kernel Parameters using Remotely Sensed Satellite Data", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, India, May 20 2016.

[10] RimahAmami, Dorra Ben Ayed, Noureddine Ellouze, "Practical Selection of SVM Supervised Parameters with Different Feature Representations for Vowel Recognition", May 9 2013, vol 7.

[11] Auria, Laura, and Rouslan A. Moro. "Support vector machines (SVM) as a technique for solvency analysis." (2008).

[12] Ms. Hetal Bhavsar, Dr. Amit Ganatra, "Radial Basis Polynomial Kernel (RBPK): A Generalized Kernel for Support Vector Machine", International Journal of Computer Science and Information Security (IJCSIS), April 2016, vol 14, no 4.

[13] Nusrat Parveen, Sadaf Zaidi, Mohammad Danish, "Support vector regression model for predicting the sorption capacity of lead (II)", June 11 2016, pp 629 – 631.

[14] Debasish Basak, Srimanta Pal and Dipak Chandra Patranabis, "Support Vector Regression", October 2007, Vol 11 No 10, pp 203-223.

[15] Xiaodan Yu, Zhiquan Qi, Yuanmeng Zhao, "Support Vector Regression for Newspaper/Magazine Sales Forecasting", International Conference of Information technology and Quantitative Management, 2013, pp 1055-1062.

[16] R. Usha Rani, Dr. T.K Rama Krishna Rao, "An Enhanced Support Vector Regression Model for Weather Forecasting", IOSR Journal of Computer Science, May-June 2013, vol 12 iss 2, pp 21-24.

[17] V Kranthi Sai Reddy, "Stock Market Prediction Using Machine Learning," International Research Journal of Engineering and Technology, e-ISSN: 2395-0056, p-ISSN: 2395-0072, vol 5 issue 10, Oct 2018, pp 1032 – 1035.

[18] Troy J. Strader, John J. Rozycki, Thomas H. Root, Yu-Hsiang (John) Huang", Machine Learning Stock Market Prediction Studies: Review and Research Directions," Journal of International Technology and Information Management, vol 28 issue 4, 2020, pp 63-84.

[19] Jingyi Shen and M. Omair Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system," Journal of Big Data, July 30, 2020.

[20] M Umer Ghani, M Awais, Muhammad Muzammul, "Stock Market Prediction Using Machine Learning(ML) Algorithms," Advances in Distributed Computing and Artificial Intelligence Journal, eISSN: 2255-2863, vol 8 issue 4, 2019, pp 97 – 116.

[21] Subhadra Kompella, Kalyana Chakravarthy Chilukuri, "Stock Market Prediction Using Machine Learning Methods," International Journal of Computer Engineering & Technology, ISSN Print: 0976-6367, vol 10 issue 3, June 2019, pp 20 – 30.

[22] Dinesh A. Kulkarni, et al., "Stock Market Prediction Using Machine Learning In Python," International Journal Of Innovative Research In Technology, ISSN: 2349-6002, vol 6 issue 3, August 2019, pp 167- 171.

[23] Nitin Nandkumar Sakhare, S. Sagar Imambi, "Performance Analysis of Regression Based Machine Learning Techniques for Prediction of Stock Market Movement," International Journal of Recent Technology and Engineering, ISSN: 2277-3878, vol 7 issue 6S4, April 2019, pp 206-213.
[24] K. Hiba Sadia, Aditya Sharma, Adarrsh Paul, SarmisthaPadhi, Saurav Sanyal, "Stock Market Prediction Using Machine Learning Algorithms," International Journal of Engineering and Advanced Technology, ISSN: 2249 – 8958, vol 8 issue 4, April 2019 pp 25-31.
[25] Nargis I Khan, Prof. Nitin Janwe, "Stock Market Prediction Using Machine Learning Approach: A Review," International Journal of Innovations in Engineering and Science, e-ISSN: 2456-3463, vol 4 issue 8, 2019, pp 272-281.
[26] Ashwini Pathak, Sakshi Pathak, "Study of Machine learning Algorithms for Stock Market Prediction," International Journal of Engineering Research & Technology, ISSN: 2278-0181, vol 9 issue 6, June 2020, pp 295-300.

Razia Sulthana A completed her Bachelor of Technology (B.Tech) in Information Technology under Anna University India and Master of Engineering in Computer Science from Anna University. She holds MBA from University of Madras. She then completed her Doctor in philosophy in Computer Science from SRM Institute of Science and Technology. She is currently working with BITS Pilani, Dubai Campus in the department of Computer Science. She has published a number of research articles in reputed journals and in conferences. Her area of interest includes, Machine Learning, Recommendation System. A number of certifications is made by her with Microsoft and Cambridge. She is the reviewer of a number of Elsevier and Springer journals.

Pranav Ramesh is pre final year Computer Science Engineering student at Birla Institute of Technology & Science Pilani, Dubai. He is a lover of mathematics and well versed in Python programming. He always strives to find the bridge between Mathematics and Computer Science. He is currently pursuing a minor in Data Science and is working forward to publish papers .

ISSN 2210-142X

Face Identification Approach using Legendre Moment and Singular Value Decomposition

Mohammed Hasan Abdulameer [1], Raaed Adnan Kareem[2]
[1] University of Kufa/ Faculty of education for women, Department of computer science, Najaf, Iraq
[2] University of Kufa/ Faculty of computer science and mathematics, Najaf, Iraq
Email: mohammed.almayali@uokufa.edu.iq, Email: raaed.79ak@gmail.com

ABSTRACT

Face recognition refers to the identification of a person based on facial features. A facial feature can be used in a variety of computer vision techniques, including face detection, expression detection, and a variety of video surveillance This paper presents a facial identification approach for dealing with face images from video cameras such as CCTV cameras. The proposed method is divided into three stages: preprocessing, feature extraction, and identification. First, the preprocessing stage relied on face detection, cropping, and unifying the image dimensions. Second, feature extraction is accomplished through the use of Legendre moment and singular value decomposition (SVD). Finally, the Manhattan classifier is used to complete the face identification. In the experiments, two face datasets are used: SCface dataset from surveillance cameras and ORL face dataset taken under diverse circumstances. The best performance for the ORL database was (98.75%) percent, whereas the best result for the SCface database was (99%) percent.

Keywords: Biometric, Face Identification, Legendre moment , SVD , Manhattan Classifier,

1. INTRODUCTION

There are many security threats that people face in their daily lives in this era, such as terrorist threats, assassinations, robberies, and so on, which necessitate the use of impactful techniques to detect and identify the perpetrators of these criminal acts. Face recognition is frequently used in the detection of thieves and intruders. This technology is not limited to tracking criminals, it may be used to find children, the elderly, and people without a missing mind or memory. This face recognition technology is an effective tool that has a wide variety of applications in the areas of information retrieval, self - service banking services, security authentication, etc. Face recognition is either identification or verification. identification, if we need to recognize the person from many people (one-to-many), either verification, it's recognized the person if he or not (one-to-one) [1]. One of the most important and useful factorizations in linear algebra is singular value decomposition (SVD). We describe how SVD is used to solve face feature extraction difficulties [2]. Besides, Legendre moments are orthogonal moments that have been used in a variety of applications of pattern recognition [3]. In this paper we propose effective method to identify the face image from many images by using Legendre moment and singular value decomposition (SVD). The remaining sections of the paper are organized as follows: related works are discussed in section 2, theoretical background about the topic is presented in section 3, proposed methodology is explained in section 4, experimental results are shown in section 5, and the conclusion is presented in section 6.

2. RELATED WORK

Face recognition has come a long way during the last years. The success of face recognition depends on the feature extraction and recognition, which have been discussed in almost all previously proposed techniques [4].

In [5], The principal component analysis (PCA) and neural networks are merged. In the database, the most prominent features of a variety of human faces were extracted, and the corresponding values were determined. As a result, if a new image was inserted into the system for recognition, the key features were extracted and the contrast between the source images and the stored images was calculated. As a result, some distinguishing characteristics of the most recent facial image will be lost.

The framework used in [6] is a hybrid of Gabor wavelets and General Discriminant Analysis (GDA), and it is referred to as appearance-based since features are extracted from the original face image. The feature vectors were subspace modeled. The development of Gabor filters for facial feature extraction is also discussed. For both identification and verification works, the technique has been thoroughly tested. Gabor waves greatly improve machine performance.

In [7], A face recognition procedure created on LBP and LNMF is proposed. According to recent neighborhood distance classification, the parameter matrix can be obtained from a test sample database containing a projection of a non-negative subspace, and better pattern recognition results were eventually achieved. The results of the experiment showed that the proposed method would dramatically improve image recognition rates.

In [8], SVD was used to represent each face image whereas Individual values are able to represent face images, and individual values of the face images at different resolutions have a nearly linear connection. The appropriate high resolution (HR) face image pairings for each input low resolution (LR) face can then be selected from the face gallery. The mapping functions to interpolate the two matrices in the SVD representation can be learned based on the selected (low resolution and high resolution) pairs to more accurately recreate HR face images. As a result, the final assessment of high-frequency detail in high-resolution facial photographs is more accurate and dependable.

Furthermore, in [9] a new method for producing virtual images from original images using single value analysis (SVD) is proposed in order to obtain richer face representations. The virtual images acquired increase the size of the training sample set while also representing reasonably stable low-frequency facial information, resulting in improved durability and classification accuracy. They integrated the virtual samples with the genuine samples, allowing them to identify the items with additional information. In this paper, the legendary moment approach with SVD is proposed as face identification solution for security purposes.

3. THEORETICAL BACKGROUND

A. Legendre moments

Teague [3] in 1980, introduced the Legendre moments, which are moments with Legendre polynomials as kernel functions. Legendre moments are orthogonal moments that have been used in a variety of applications of pattern recognition. They can be used to achieve a near-zero continuity measure in a series of moment functions, allowing the moments to correspond to the image's independent characteristics. The projection of the image intensity function into Legendre polynomials is used to define Legendre moments. With image intensity function f(x,y), the two-dimensional Legendre moments of order (a+b) are described as follows [10][11]:

$$LP_{ab}=\frac{(2a+1)\times(2b+1)}{4}\int_{-1}^{1}\int_{-1}^{1}P_a(x)P_b(y)f(x,y)dxdy \quad (1)$$

x and y belong to the interval [-1,1]

Legendre polynomial, $LPa(x)$, of order a is described as follows:

$$LPa(x)=\sum_{k=0}^{a}\left\{(-1)^{\frac{a-k}{2}}\frac{1}{i}\frac{(a+k)!x^k}{\left(\frac{i-k}{2}\right)!\left(\frac{i+k}{2}\right)!k!}\right\}_{a-k=\text{ even}} \quad (2)$$

Legendre polynomials' recurrence relation, $LPa(x)$ become as follows:

$$LPa(x)=\frac{(2a-1)xLP_{a-1}(x)-(a-1)LP_{a-2}(x)}{a} \quad (3)$$

whereas $LP_0(x)=1, LP_1(x)=x$ and $i>1$. Whereas, the space defined for the Legendre polynomial is within an interval $[-1,1]$, an equal dimensions image (N ×N) pixel with $f(i,j)$ intensity function. $0\leqslant i,j\leqslant(N-1)$ is scaled in the space of $-1<x,y<1$. In the result of this, (1) can be expressed in discrete form as:

$$LP_{ab}=\lambda_{ab}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}P_a(x_i)P_b\left(y_j\right)f(i,j) \quad (4)$$

whereas the normalizing constant,

$$\lambda_{ab}=\frac{(2a+1)(2b+1)}{N^2} \quad (5)$$

The normalized pixel coordinates in the interval [-1, 1] are represented by x_i and y_j, that are given in (6) and (7)

$$x_i=\frac{2i}{N-1}-1 \quad (6)$$

$$y_j=\frac{2j}{N-1}-1 \quad (7)$$

B. Singular Value Decomposition (SVD)

Singular Value Analysis (SVD) is the foundation for a variety of other techniques, such as classification techniques, dynamic mode decomposition (DMD), and proper orthogonal decomposition (POD). When it comes to manipulating data from complex structures, high dimensions are a common problem. Large size data sets, such as audio, image, or video data, can be included in these systems. Images are components of a high-dimensional vector space since they usually involve a large number of measurements (pixels). Many images, on the other hand, are compressible, which means that related information can be expressed in a much smaller subspace. In terms of dominant patterns, SVD provides a systematic method for determining a low-dimensional approximation of high-dimensional results. Data-driven technology that detects patterns solely through data, without the use of professional experience or intuition. The SVD algorithm is numerically stable and provides a hierarchical representation of data in terms of a new coordinate system described by the data's dominant correlations. In addition, unlike the eigen-decomposition setup, the SVD is guaranteed to have some matrix. To analyze a large set of data such as $\mathbf{Z}\in\mathbb{C}^{n\times m}$:

$$Z=\begin{bmatrix}| & | & & | \\ z_1 & z_2 & \cdots & z_m \\ | & | & & |\end{bmatrix} \quad (8)$$

Images transformed into column vectors with a number of elements such as pixels in the image are represented $z_k \in \mathbb{C}^n$ columns. The k-index is a mark that denotes the set of k-characteristic measures. Z consists of a time-series of facts, and $\mathbf{z}_k = \mathbf{z}(k\Delta t)$. The n-state dimension is often very large, in the order of millions or more of degrees of freedom. Snapshots are also referred to as columns, and m represents the number of snapshots in Z. For several systems $n \gg m$, it results in a tall matrix, in contrast to a short matrix at $n \ll m$. Any matrix with a complex value $\mathbf{Z} \in \mathbb{C}^{n\times m}$ has a unique matrix analysis called SVD:

$$Z = U\Sigma V^* \qquad (9)$$

$\mathbf{U} \in \mathbb{C}^{n\times n}$ is unitary matrix, with orthonormal columns, also $\mathbf{V} \in \mathbb{C}^{m\times m}$. but $\Sigma \in \mathbb{R}^{n\times m}$ is a matrix with real. positive or zeros entries on the diagonal and zeros off the diagonal. The simple (*) means the transpose of a complex conjugate. U and V must be unitary to be used widely.

When n $\geq$ m, On the diagonal, the matrix s has at most m positive or negative components, and it could be written as: $\Sigma = \begin{bmatrix}\hat{\Sigma}\\0\end{bmatrix}$.

The economy SVD can be used to precisely describe Z [12]:

$$Z=U\Sigma V^*=[\hat{U} \quad \&\&\hat{U}^{\perp}]\begin{bmatrix}\hat{\Sigma}\\0\end{bmatrix}V^*=\hat{U}\hat{\Sigma}V^* \qquad (10)$$

The full SVD and economy SVD are shown in Fig. 1 and Fig. 2.

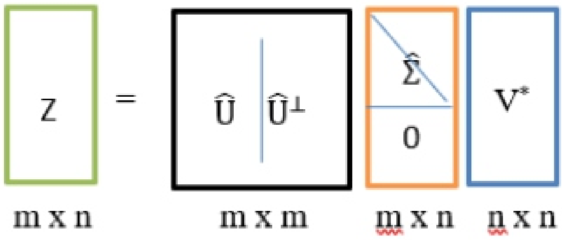


Figure 1. Full SVD

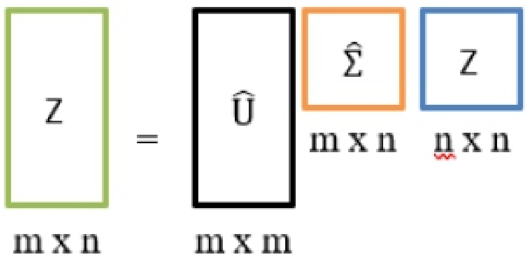


Figure 2. Economy SVD

All columns of $\hat{U}^{\perp}$ cover vectors that are matching and orthogonal with that spanned by $\hat{U}$. The left singular vectors of Z are the columns of U, and the right singular vectors are the columns of V. Singular values are the diagonal elements of $\hat{\Sigma} \in \mathbb{C}^{m\times m}$, and they are sorted from largest to smallest. The number of non-zero singular values determine Z's rank. The numerical implementation of SVD is both essential and mathematically advantageous, as it is the backbone of computing science and engineering [13].

C. Manhattan Classifier

The Manhattan distance [14], which is used to calculate the distance between two points or vectors, is often used in the classification and identification of faces in image processing and computer vision. The Manhattan distance is the number of two points or vectors' differences. The Manhattan distance between point $a = (a_1, a_2, a_3 \dots a_n)$ and point $b = (b_1, b_2, b_3 \dots b_n)$ is:

$$Manhtn(A,B)=\sum_{i=1}^{n} |a_i - b_i| \qquad (11)$$

4. THE PROPOSED METHODOLOGY

In this study, we propose a new face recognition system based on a combination of face features from two approaches by combining Legendre moments with the SVD algorithm and the Manhattan Distance Classifier is for face identification. Fig. 3 depicts the whole proposed method:

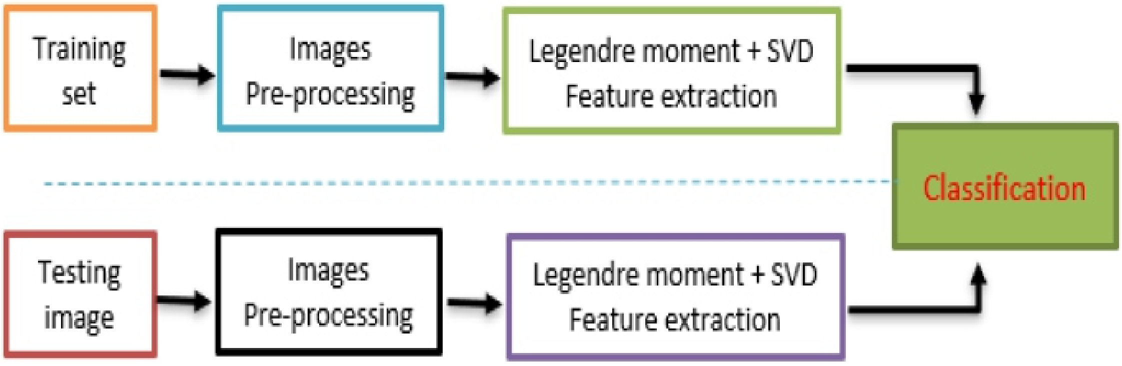


Figure 3. The Proposed Methodology

The first step is to find Legendre polynomial which is utilized to find the Legendre moments. Obtaining a Legendre polynomial is quite different from finding a Legendre moment that has nothing to do with the picture. At the beginning, we take the image rows and columns' dimensions, the lowest of which is the Legendre polynomial length. The number of rows in this Legendre polynomial is determined by the Maximum order, denoted by m, and the number of columns in the Legendre polynomial. the Legendre polynomial represents the shortest distance between rows or columns. it is based on three steps as follow:

- The first row in polynomial will be ones only as formula (12):

$$P(1,i+1) = 1 \qquad (12)$$

- The second row in polynomial is calculated according to (6):
- The number of remaining rows depends on the previous rows that were calculated on (12) and (6), which in turn depend on maximum order [3].

After completing the Legendre Polynomial calculation, the Legendre moments calculation begins, where two rows are taken from the Legendre polynomial matrix and multiplied together as in (4). The result is dotted with the image data, then we find the sum for this matrix and then we multiply the result in (5) to get the value of the Legendre moment. Programmatically, by applying the Legendre moment function with a value of maximum order = 5 to the image, we get an array of features with a size of 37 x 37, converting this array into a features vector RX.

For the SVD algorithm: a set of steps is applied to the image to obtain two parameters: Eigen value and Eigen vector, where we calculate the SVD for each 5 x 5 sub-array of the image to extract the Eigen value and Eigen vector and then put these values into a uniform vector at the end. We can summarize the entire process of adopting SVD by going through the following Equations:

Compute the square matrix C1 by (13), since, Z is a 5x5 sub-array:

$$C1 = Z^T \times Z \qquad (13)$$

Find the eigen values λ_i from the (14), Since, I mean a determent matrix and S represent a matrix of eigen values.

$$| C1 - \lambda_I \times I | = 0 \qquad (14)$$

Since,

$$S = \begin{bmatrix} \lambda_{1,1} & \cdots & \lambda_{1,5} \\ \vdots & \ddots & \vdots \\ \lambda_{5,1} & \cdots & \lambda_{5,5} \end{bmatrix} \qquad (15)$$

After that, we need to extract U matrix based on (16), to get the eigen vectors

$$C2 = Z \times Z^T \qquad (16)$$

Then we compute the eigen values (λ_i) of C2 matrix based on (17):

$$| C2 - \lambda_i \times I | = 0 \qquad (17)$$

And we find the corresponding eigen vectors (r_i)based on (18), (19) and (20):

$$(C2 - \lambda_i \times I) \times r_I = 0 \qquad (18)$$

r_i must be orthonormal.

$$u_i = \frac{r_i}{\|r_i\|} \qquad (19)$$

Since, u_i represent values of the eigen vectors whereas U represent a matrix of eigen vectors.

$$U = \begin{bmatrix} u_{1,1} & \cdots & u_{1,5} \\ \vdots & \ddots & \vdots \\ u_{5,1} & \cdots & u_{5,5} \end{bmatrix} \qquad (20)$$

The SVD implements on 5x5 sub-array, this sub-array passes over all of the pixels of the image and extracts four features from it to put in the vector D; these four features are the value of the position (1,1) of the U array and the values of the positions (1,1), (2,2), and (3,3) from the matrix S, as shown in (21):

$$D = [U(1,1)\ S(1,1)\ S(2,2)\ S(3,3)] \qquad (21)$$

Then, in the matrix XT, the features of vector D are combined with the features of vector RX, as shown in (22):

$$XT = [RX\ D] \qquad (22)$$

In general, the SVD produces the greatest number of features for that image because there are three matrices that are close to the original image, so the number of features produced by the SVD is greater than the number of image vectors. Each row of the matrix XT represents a single feature vector of the training images obtained from D and RX. Similarly, the image of the testing process is managed in the same way that the image of the training process is managed. The image is then classified using the Manhattan Classifier to determine the identity of the intended person, whether he is present or not, using (11).

5. EXPERIMENTAL RESULTS

A. Experimental Results of Orl Dataset

This database contains images of 40 participants [15], each participant has ten different images, each with a different facial expression and angle. Fig. 4 show samples from this dataset. We run the proposed algorithm (Legendary + SVD) five times with five different cases: 90% training vs 10% testing, 80% training vs 20% testing, 70% training vs 30% testing, 60% training vs 40% testing, and 50% training vs 50% testing. We discovered that the highest accuracy rate was 98.75 when the training rate was 80% and the test rate was 20%, while the lowest accuracy was 94 when both the training and testing rates were 50%. Table II and Fig. 5 analysis of the results obtained of the cases we utilized.

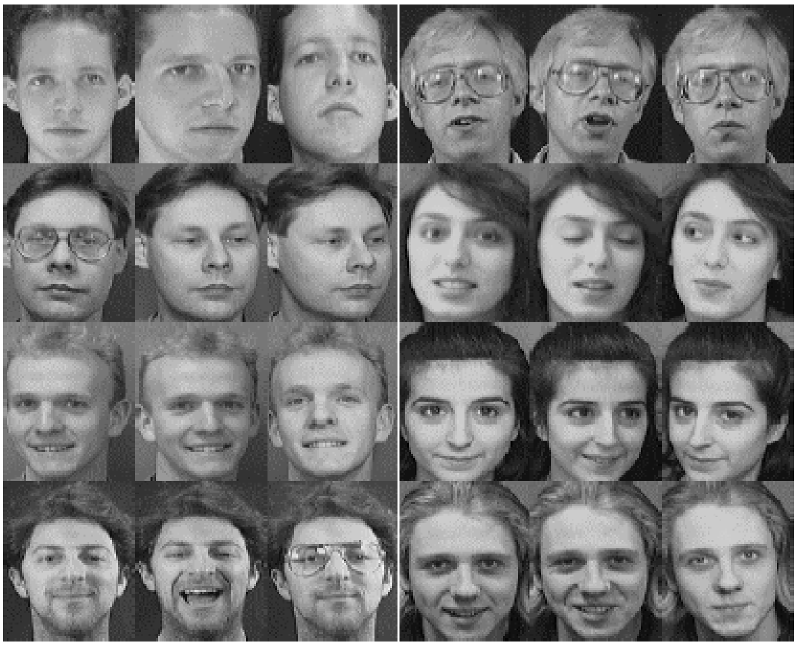


Figure 4. Samples from ORL dataset

TABLE I. RECOGNITION RATES OF THE PROPOSED LEGENDARY & SVD FOR ORL DATASET

K-Fold	No. of Training Images	No. of Testing Images	Recognition Accuracy
1	360	40	95
2	320	80	98.75
3	280	120	97.5
4	240	160	96.875
5	200	200	94

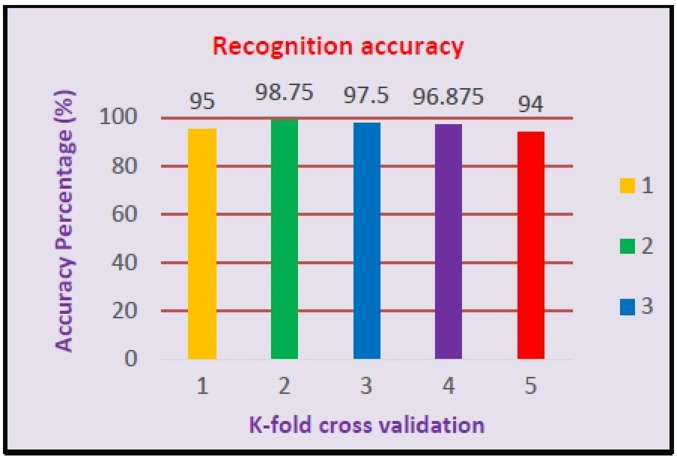


Figure 5. Recognition Rates of the Legendary & SVD for ORL Dataset

Furthermore, we compared the proposed Legendary + SVD method to legendry and SVD separately to demonstrate the method's stability. all the methods are tested on ORL dataset, with a training rate of 80% training and 20% testing that yielded the best results previously. The obtained results showed that the proposed Legendary + SVD had higher accuracy than Legendary and SVD, as shown in Table II and Fig. 6.

TABLE II. COMPARATIVE STUDY OF RECOGNITION RATE OF LEGENDRE, SVD AND LEGENDARY + SVD ON ORL DATASET

K-fold	Recognition Accuracy		
	Legendre Moments	SVD	Legendre + SVD
1	86.25	93.75	97.5
2	83.75	95	98.75
3	85	96.25	100
4	86.25	95	97.5
5	81.25	92.5	98.75

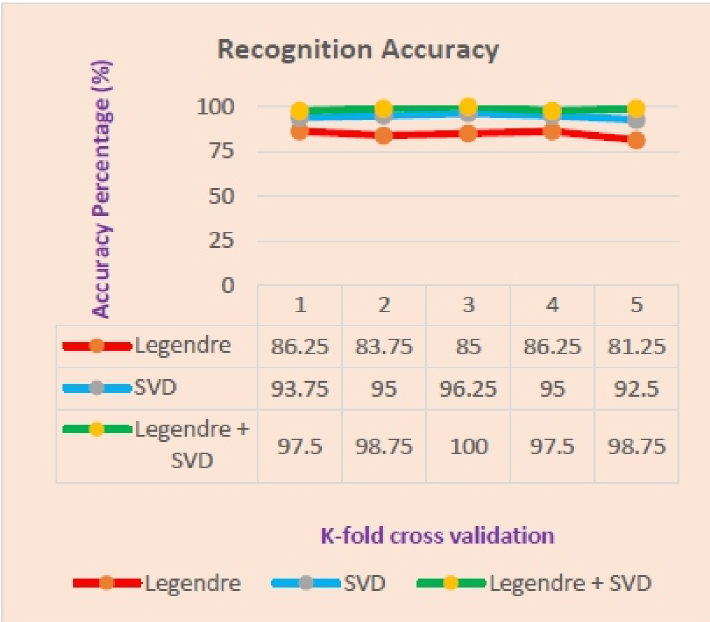


Figure 6. Comparative study of recognition rate of Legendre, SVD and legendary + SVD on ORL dataset

Experimental Results Of Scface Dataset

In this database [16], 130 people were photographed using five different resolution video surveillance cameras (cam1, cam2, cam3, cam4, and cam5). Cam1 and Cam5 were used for night monitoring where the images taken from cam1 at night assigned as cam6, and images taken by cam5 at night identified as cam7. The following procedures are followed by all database participants: Each person stands in front of the surveillance cameras in both the light and the dark. the distances were (4.20 meters, 2.60 meters, and 1 meter), each person had 21 photos taken in various lighting conditions and at various distances, directions, and angles.

• The Results of Distance-1

Seven cases were tested at the first distance (4.20 meters), with images taken from Cam2-Cam7 for training and Cam1 for testing. The sample images from the chosen distance are depicted in Fig. 7. The accuracy of face recognition was low at the chosen distance due to a lack of image resolution, long shooting distance, different lighting conditions and shooting angles for each camera. The test images taken from cam2 and cam4 achieved the best recognition of 90% in this distance, while the test images taken from cam6 with the lowest resolution achieved the lowest recognition of 25%. Before the initial processing, the dimensions of the original images in this distance were 100 x 75 pixels, which were unified to 90 x 70 pixels and the face is detected and cropped for every image. Table III and Fig. 8 shows the details of the results for distance 1 with 7-fold cross validation.

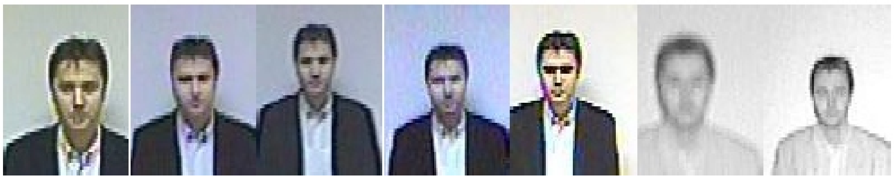


Figure 7. Samples of 7 Cameras from Distance-1

TABLE III. RECOGNITION RATES FOR DISTANCE-1, (4.20 METERS) BASED ON LEGENDRE MOMENTS AND SVD

K-fold	Training images	Testing images	Recognition accuracy
1	480 (Cam2,3,4,5,6,7)	80-Cam1	87.5
2	480 (Cam1,3,4,5,6,7)	80-Cam2	90
3	480 (Cam1,2,4,5,6,7)	80-Cam3	78.75
4	480 (Cam1,2,3,5,6,7)	80-Cam4	90
5	480 (Cam1,2,3,4,6,7)	80-Cam5	81.25
6	480 (Cam1,2,3,4,5,7)	80-Cam6	25
7	480 (Cam1,2,3,4,5,6)	80-Cam7	32.5

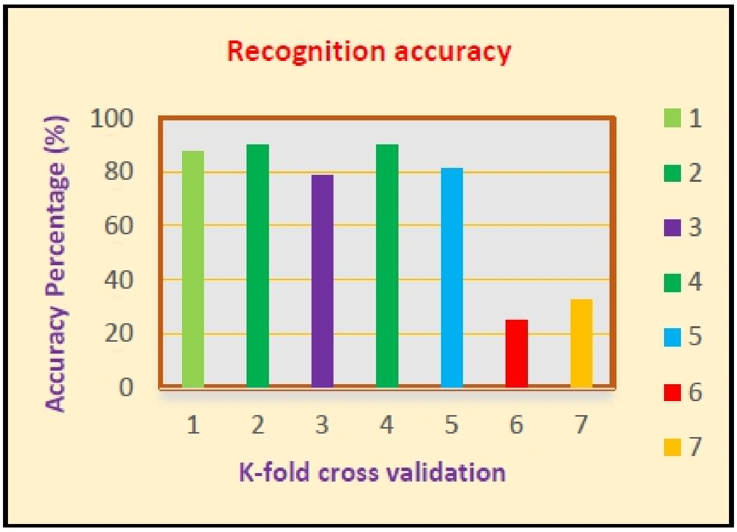


Figure 8. Recognition Rates for distance-1

- **The Results of Distance-2**

In order to efficiently examine the proposed method, seven cameras are used in this test, but at different distances of 2.60 meters. The images in this dimension had original dimensions of 144 x 108, but after preprocessing, the dimensions became 100 x 85. We utilized 560 images, including 480 training images and 80 testing images. Fig. 9 depicts different samples of images from seven cameras and the proposed method is implemented using k-fold cross validation. The best recognition accuracy was 98.75 percent for test images taken from cam2 and cam3, while the worst recognition accuracy was for cam6 test images, as shown in Table IV and Fig. 10.

Figure 9. Samples of 7 Cameras from Distance-2

TABLE IV: RECOGNITION RATES FOR DISTANCE-2 (2.60 METERS) BASED ON LEGENDRE MOMENTS AND SVD

K-fold	Training images	Testing images	Recognition accuracy
1	480 (Cam2,3,4,5,6,7)	80-Cam1	96.25
2	480 (Cam1,3,4,5,6,7)	80-Cam2	98.75
3	480 (Cam1,2,4,5,6,7)	80-Cam3	98.75
4	480 (Cam1,2,3,5,6,7)	80-Cam4	96.25
5	480 (Cam1,2,3,4,6,7)	80-Cam5	55
6	480 (Cam1,2,3,4,5,7)	80-Cam6	40
7	480 (Cam1,2,3,4,5,6)	80-Cam7	43.75

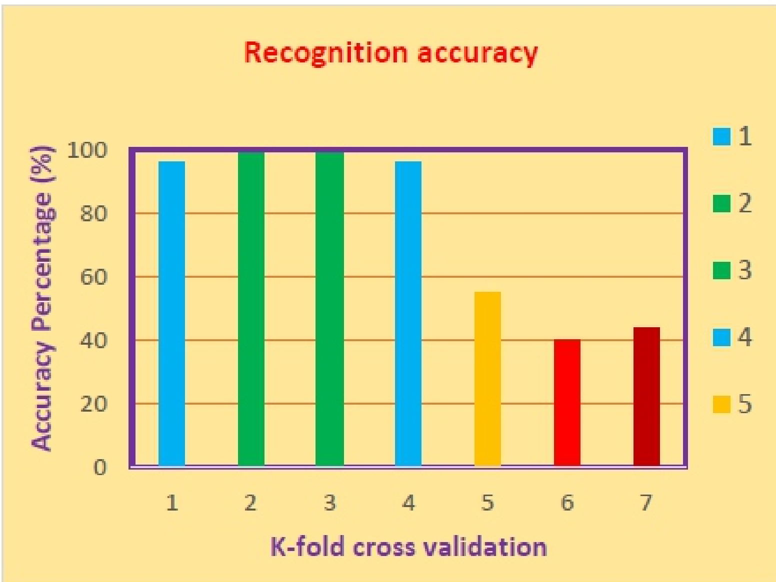


Figure 12. Recognition Rates for distance-3

• **The Results of Combining All Three Distances**

In this case, 500 images were used at random from the cameras for the three dimensions, whereas 10 images were collected from the three dimensions for each person, as seen in Fig. 13. The 5-fold cross validation was used in the test. The best result of the proposed method was (99%) as shown in the table VI and Fig. 14.

Figure 13. Samples from Three Distances

TABLE VI. RECOGNITION RATES FOR COMBINING ALL THREE DISTANCES

K-fold	No. of training images (80%)	No. of testing images (20%)	Recognition accuracy
1	400	100	98
2	400	100	99
3	400	100	97.27
4	400	100	95.45
5	400	100	78.18

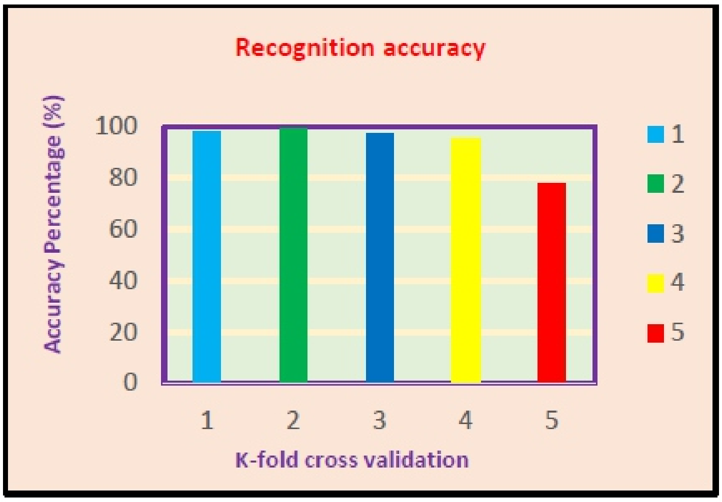


Figure 14. Recognition Rates for combining all three distances

In Table VII and Fig. 15. we made a comparison of the proposed algorithm (Legendre + SVD) with both Legendre and SVD algorithm, where we used 240 images from each distance in SCface database, (240 (80%) for training and 40 (20%) for testing for each of the three algorithms. The best recognition accuracy in our proposed algorithm was 97.5% in third distance, while the best recognition accuracy in

Legendre was 77.5% in second distance, and in the SVD algorithm, the best recognition accuracy was 92.5% in third distance, and thus it becomes clear that our proposed algorithm is superior to both Legendre and SVD.

TABLE VII. COMPARATIVE STUDY OF RECOGNITION RATE OF LEGENDRE, SVD AND LEGENDRE + SVD ON SCFACE DATASET

Distance	Recognition Accuracy		
	Legendre Moments	SVD	Legendre + SVD
D1	47.5	80	87.5
D2	77.5	90	95
D3	55	92.5	97.5

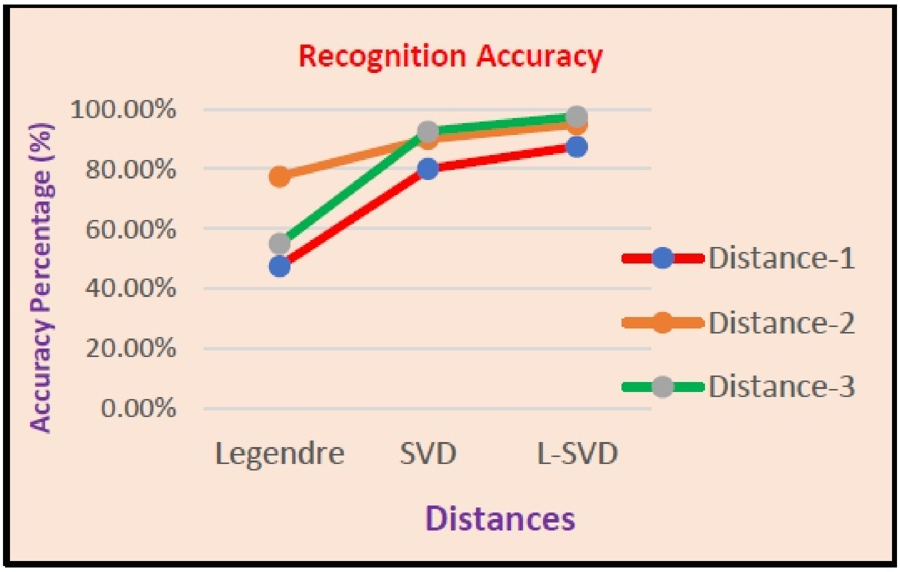


Figure 15. Comparative Study of Recognition Rate of Legendre, SVD and Legendre + SVD on Scface Dataset

6. CONCLUSION

This paper describes a method for recognizing faces that is based on Legendre moments and singular value decomposition (SVD). The Legendre moments and SVD are used for facial feature extraction, while the Manhattan classifier is used for recognition. The ORL dataset and the SCface dataset were both used in the study. We demonstrated in both datasets that Legendre with SVD is sufficient for face recognition tasks in a variety of situations. The power of the proposed method is demonstrated by the fact that it can produce consistent matching results on images from various scenarios. Finally, we believe that using more powerful classifiers would significantly improve the current performance of face recognition. The proposed method can be used in real-world applications such as law enforcement agencies, airports, and building access management, among others. Future research will look into and investigate the use of a variety of classifiers that can be assigned based on the input face's class.

7. REFERENCES

[1] H. Moon, "Biometrics person authentication using projection-based face recognition system in verification scenario," in International Conference on Biometric Authentication, 2004, pp. 207–213.

[2] N. Muller, L. Magaia, and B. M. Herbst, "Singular value decomposition, eigenfaces, and 3D reconstructions," SIAM Rev., vol. 46, no. 3, pp. 518–545, 2004.

[3] M. R. Teague, "Image analysis via the general theory of moments," JOSA, vol. 70, no. 8, pp. 920–930, 1980.

[4] M. H. Abdulameer, S. N. H. Sheikh Abdullah, and Z. A. Othman, "Support vector machine based on adaptive acceleration particle swarm optimization," Sci. World J., vol. 2014, 2014.

[5] A. S. S. Navaz, T. Dhevisri, and P. Mazumder, "Face recognition using principal component analysis and neural networks," March-2013, Int. J. Comput. Networking, Wirel. Mob. Commun. Vol, no. 3, pp. 245–256, 2013.

[6] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," Image Vis. Comput., vol. 25, no. 5, pp. 553–563, 2007.

[7] M. Wu and T. Lu, "Face Recognition based on LBP and LNMF Algorithm," in 2016 15th International Symposium on Parallel and Distributed Computing (ISPDC), 2016, pp. 368–371.

[8] M. Jian and K.-M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 11, pp. 1761–1772, 2015.

[9] G. Zhang, W. Zou, X. Zhang, and Y. Zhao, "Singular value decomposition based virtual representation for face recognition," Multimed. Tools Appl., vol. 77, no. 6, pp. 7171–7186, 2018.

[10] S. Annadurai and A. Saradha, "Face Recognition Using Legendre Moments.," in ICVGIP, 2004, pp. 461–466.

[11] C.-W. Chong, P. Raveendran, and R. Mukundan, "Translation and scale invariants of Legendre moments," Pattern Recognit., vol. 37, no. 1, pp. 119–129, 2004.

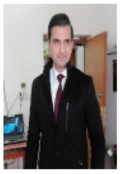
[12] A. A. Aljarrah and A. H. Ali, "Human Activity Recognition by Deep Convolution Neural Networks and Principal Component Analysis," Furth. Adv. Internet Things Biomed. Cyber Phys. Syst., p. 111, 2021.

[13] N. K. E. L. Abbadi and E. Saleem, "Automatic image colorization based on SVD and lab color space," 2018.

[14] K. M. Ponnmoli and D. S. Selvamuthukumaran, "Analysis of Face Recognition using Manhattan Distance Algorithm with Image Segmentation," Int. J. Comput. Sci. Mob. Comput., vol. 3, no. 7, pp. 18–27, 2014.

[15] A. L. and Cambridge, "The ORL Database of Faces." https://www.researchgate.net/publication/221786184_PCA_and_LDA_Based_Neural_Networks_for_Human_Face_Recognition.

[16] M. Grgic, K. Delac, and S. Grgic, "SCface–surveillance cameras face database," Multimed. Tools Appl., vol. 51, no. 3, pp. 863–879, 2011.



Mohammed Hasan Abdulameer received his BSc in computer science from in 2002 from Al mamoun university, MSc in computer science in 2006 from Iraqi commission form computer and informatics and PhD computer science from national university of Malaysia in 2015. He is currently assistant professor in Kufa University, Iraq. His fields of interest are artificial intelligence, Natural language processing, face recognition.

Raaed Adnan Kareem is currently a M.Sc. student in faculty of computer science and mathematics at Kufa University, Najaf, Iraq. His research areas include computer vision and Artificial intelligence.

ISSN 2210-142X

True Feature based Fashion Recommendation with Customer Reviews using CNN and LSTM

Zafar Ali Khan N[1*], R Mahalakshmi[2]
[1]Research Scholar, Department of CSE, School of Engineering, Presidency University, Bengaluru, Karnataka, India
[2]Associate Professor and Program Chair CSE, Department of CSE, School of Engineering, Presidency University, Bengaluru, Karnataka, India
*Corresponding Author: zafaralikhan@presidencyuniversity.in, Tel.: +00-917338021015

ABSTRACT

The fast development of fashion through internet based business, design suggestion has becomes primarly get advanced. It is an showcasing tools based on clients and appraisals. The online fashion bussines is a great hotspot for understanding clients' shopping and encounters, item/thing exhibitions. This e-commerce platform subsequently providing best valuables for improving customized suggestions for future buys. Nonetheless, most of Fashon recomemdation reviews proposal techniques need effective systems to incorporate neighbourhood and worldwide clients' appraisals and surveys. This paper proposed the feature-based fashion recommendation approach with attention mechanism (FFRAM) to foresee client appraisals dependent on online survey design items. This model can extricate inactive angle highlights about clients and things independently through two similar ways of convolutional-neural-networks (CNN), long short term memory (LSTM), and consideration systems. One way measures client surveys and different adapts to thing Reviews. On every way, CNN and LSTM are combined with a consideration instrument to catch nearby perspective highlights. The worldwide angle includes separately, which are joined through an everyday activity module. The shared procedure on the two ways can improve the speculation of the FFRAM model. The extricated highlights from the two methods are additionally converged to anticipate clients' appraisals. Proper client surveys and assessments gathered from two eminent business sites were utilized to prepare and test FFRAM. The trial results exhibit that FFRAM is more successful in client rating expectations when contrasted with a few best in class design recommenders.

Keywords: FFRAM, attention technique, CNN, LSTM.

1. INTRODUCTION

This paper proposes an FFRAM in light of clients' Reviews and appraisals to target neighbourhood and worldwide angle representation. This model is built with two freeways to handle client/thing review all the while, and every way has a CNN, LSTM with consideration component to catch nearby perspective highlights, and worldwide angle includes independently. To improve the FFRAM model's speculation, the neighbourhood and worldwide perspective highlights from both client and thing Reviews are converged through everyday activities preceding the rating forecast. There are two datasets, shoes, a dataset of jewellery, and Clothing from Amazon 5-center and surveys of other public online-retailer, have utilized for preparing& FFRAM has been tested. Moreover,the significance of removing clients' inclinations aimed at various parts of style & adequacy of our consideration modules in FFRAM have inspected rather than a few cutting edge design recommenders. The fundamental commitments of this paper might be expressed as follows:

1) another design proposal receives neighbourhood and worldwide angles to get familiar with the pertinent dormant semantic data and show the cooperation among client and thing Reviews to upgrade suggestion interpretability.
2) With viewpoint portrayal and consideration system, clients' inclinations in various design perspectives can be used to evaluate the neighbourhood and worldwide significance of every word in an Review.
3) By preparing with this current practical dataset from 2 online vendors, the FFRAM might beat some other devise recommendation approaches to predict clients' assessments.

2. REVIEW OF LITERATURE

This contributionhas been identified with research in 4 zones: surveys depending proposal, angle depending suggestion, consideration instrument, and style suggestion. Late advances in every one of these regions are presented underneath.

2.1 REVIEWS-BASED RECOMMENDATION

Many manuscripts have exhibited techniques to use client Reviews for enhancing the correctness and recommendations interpretability. For tackling the sparsity of information and problems of cold-start in customary isolating computations, some research utilized the subject exhibiting schemes, for instance, LDA (latent Dirichlet association) [5] or LDA type of analyses as in [6] –[8] for eradicating inactive semantic information from Reviews and integrated it with assessments for making suggestions more interpretable and precise. The work [9] projected two isolate FEATURE learning methods by exploiting the evaluating constant and clients text consistency and things, and approaches the two views to estimate rating. However, these investigations overlooked request of word and qualities of sentences in Review, oriented information and context missing. The significant learning schemes have been implemented to suggest the architectures relied on research with an incredible. The work [10] [11] presents that CNN, RNN [12] [13], as well as capsule-network as in [14], have been utilized extensively for eradicating semantic data logically [10] [15] through making companies became well-recognized with significant element Reviews representation and possibility architecture factorization for estimation rating. The work [16] [17] presents the word vector approach and CNN have been used to attain effectiveness with clients' practices and credits of things. The work [18] shows that RNN has been combined by factorization machine by term regularization for estimating thing rating through idle variables attained from RNN. The result [15] presents that, the prediction instance engineering with new directing is projected to discriminate the unit and depend on the user estimating rating level.

As referenced in above approach, usually, the RNN has been used for a recommendation as in [19]-[23], text clustering [25] and opinion investigation [24], which might catch global long term circumstances and semantics short sentences through cooperated cycle relations among units. However, it has been estimated that RNN is having detonating problems or evaporating inclination [26] [27]. The individual RNN known as LSTM has been formed for defeating problems in NLP for current years as in [20] [23] [28] [29]. Also, LSTM has been deployed by clustering data adaptability. Their layer output at every position could be used in the form of associated implanting for recent word as in [30].

Furthermore, the projected feed-forward company in the space of design for generating paper implanting since the research of LSTM for anticipating vector for each user reliance on customer's former buy clustering as in [28]. The LSTM has not been sufficient for learning long-term circumstances deprived of data set as in [25]. This projected bi-LSTM approach as in [31] that has selection for attaining on former and further data angle setting and attaining word statement in the overall sentence.

3. FEATURE-BASED RECOMMENDATION

Albeit did a survey put together suggestion is based on the comprehension of why clients made their evaluations, it comes up short on the capacity to catch client inclinations and thing properties, yet also powerful and fine-grained connections among clients things. This way, viewpoint-based suggestions got alluring. A viewpoint-based recommender framework means to separate angles from text-based surveys, which can be, for the most part, isolated into two gatherings. The principal bunch depends on external NLP devices to break down Review substance to learn estimation investigation perspectives [32]-[34]. For example, utilizing fine-grained viewpoint level assessment examination can naturally find the most significant perspective to upgrade future client experience [35]. The subsequent gathering, for example, AARM [36], LDA [5], [32], ASCF [38], &FLAME [37], fabricates an interior design or structure to address various viewpoints in a client or thing survey. The work [4] consolidated client and thing angle level portrayals with viewpoint significance, and afterwards assessed a general rating for any client thing pair. This contribution planned an angle mindful subject model using multinomial dissemination over the Reviews to learn distinctive client and thing viewpoints in the point space [39] [15].

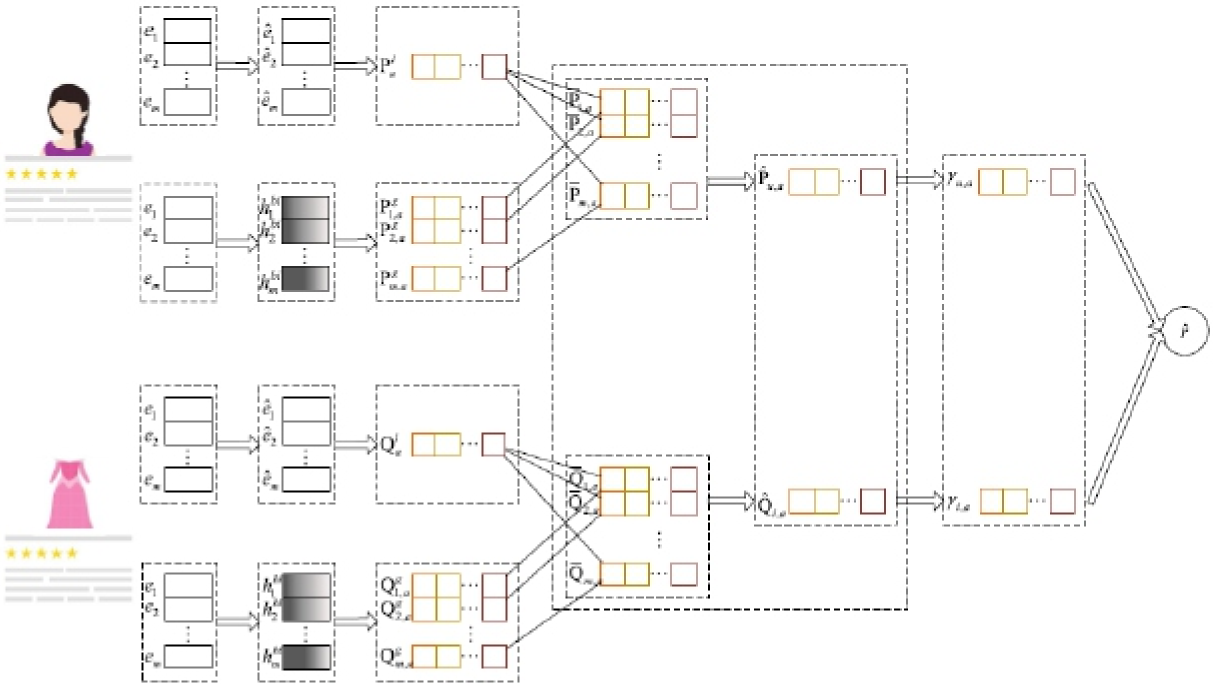


Figure 1: The framework of fashion recommendation based on the FEATURE with attention technique by utilizing item and user reviews.

3.1 ATTENTION MECHANISM

These days, consideration has been one of the influential ideas in the profound learning eld. As a vector, it is regularly the yields of thick layers utilizing the softmax work. Consideration system was first acquainted with the NLP field in [40] and still keeps up its liveliness in this field since it can deftly choose a reference for setting data to encourage worldwide learning [41]. Also, consideration component approaches incorporate threeintroductory classes. The primary type is a solitary consideration model dictated by printed level communications amid various substance pieces [42].

Moreover, subsequent classification, for example, progressive consideration [21], [43], [44] and co-consideration (neighbourhood and worldwide [2], hard and delicate [41], [45]-[47]), is a crossbreed consideration model which depends on CNN, RNN and different organizations at various levels. In [43] and [44], the word-level and sentence-level progressive consideration models have been utilized, and the yield of the main consideration model was the contribution of the subsequent consideration model. A double attending (D-Attn) model contains lenient attending (L-Attn) and global attending (G-Attn) models in equal. L-Attn plans to discover important catchphrases in the worldwide sliding window of a client/thing survey. At the same time, G-Attn intends to catch the client's general assessment articulation or thing Review. The third classification is an information-based consideration model like the progressive consideration model aside from that its information is begun from the data of other field information [48], [49]. Other consideration dependent work incorporates interpretation errands [50] and text arrangement [51].

3.2 FASHION BASED REVIEWS

Assorted style perspectives can reflect clients' inclinations in design in their Reviews. In this manner, it is exceptionally fundamental to investigate the design parts that clients are intrigued. The vast majority of the current style suggestions depend on visual pictures of apparel, shoppers' buy practices, deal information, and other significant data [52]-[60]. In any case, scarcely any investigations on design suggestion have paid enough consideration to clients' surveys and evaluations chiefly because they are not effectively interpretable. This paper will utilize surveys and assessments from the Amazon 5-center information base and another online retailer to construct a style suggestion model.

3. FEATURE-BASED FASHION RECOMMENDATION WITH ATTENTION MECHANISM (FFRAM)

Different style perspectives in their surveys can reflect clients' inclinations in design. Subsequently, it is exceptionally fundamental to investigate the parts of the design that clients are intrigued. Many current style proposals depend on visual pictures of attire, customers' buy practices, deal information, and other pertinent data [52]-[60]. Nonetheless, not many

investigations on design suggestion have paid enough considerations to clients' Reviews and evaluations fundamentally in light of the fact.

CONTEXT EMBEDDING MODULE

Estimate that consumer text the review is $R_U = r_1, r_2, \ldots, r_m$, where the notation m indicates to overall words in review as well as r_i indicates ith review word. Primarily map every word towards their embedding depiction $EU = (e_1; e_2; \ldots, e_m) \in \mathbb{R}^{m \times d}$ through the embedding context layer, where the notation e_i has been a vector of embedding for ith word. Further, the notation d indicates the number of dimensions aimed at every word embedding vector. This layer of embedding could also be regarded simply in the form of operation in embedding matrix shared that might be started by utilizing vectors name, which has pre-determined massive corpora as in [61].

A. LOCAL FEATURE FEATURE EXTRACTION MODULE

Given an implanted client/thing Review portrayal, this module's objective is to separate a bunch of nearby prospective client/thing highlights. Figure 2 delineates the interaction of nearby viewpoint highlight extraction. We can first encode the installed setting from the gigantic given survey writings and appraisals [4]. Afterwards, we use CNN to gain proficiency with the significance of each word and its viewpoint portrayal. Finally, we can get the word angle portrayal and concentrate dormant neighbourhood viewpoint semantic highlights by consideration instrument.

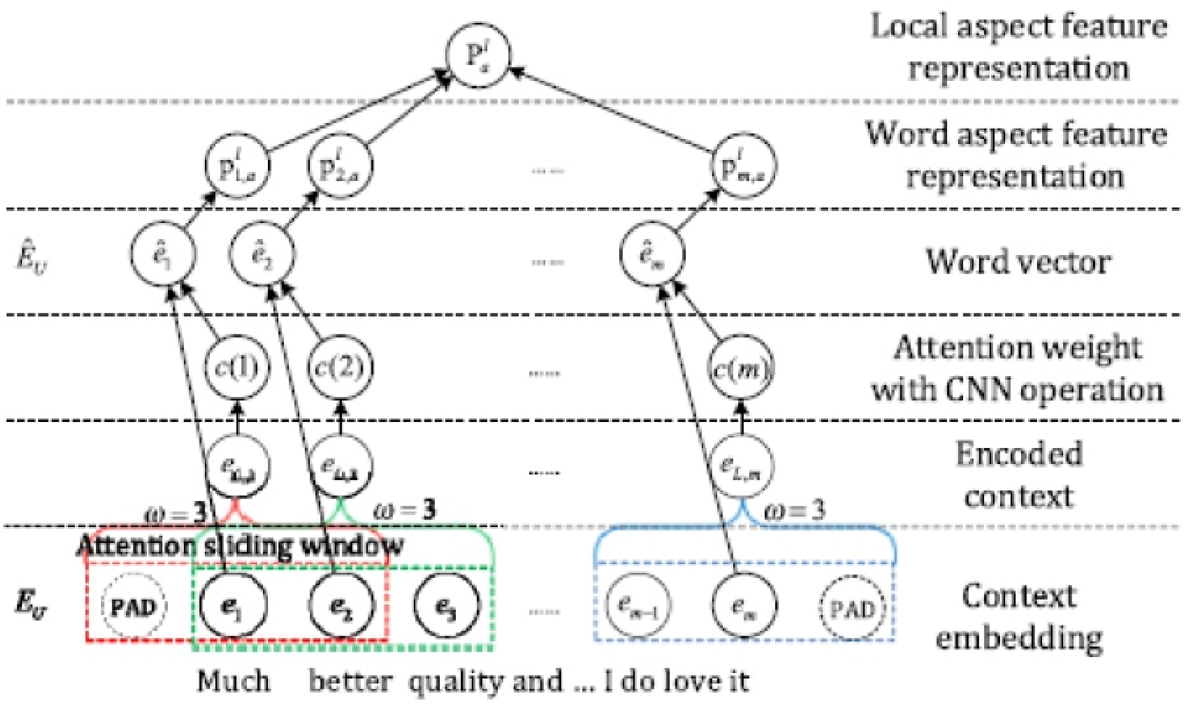


FIGURE 2. Local parameter feature mining module.

In item or user review sentence for the optimal price, the colour has been more off-white, words associated with FEATURE and emotional sentences generally happen close over each. Mainly, the significance of the ithterm in review relies on both their surrounding and name itself. One might indicate the importance of word by observing at their surrounding-words. In fig 2, the

sliding window attention with ω width has been ith word placed and later slide all over their E_U embedding vector for learning the significance of word as in [4]. Mainly, the context review has been encoded by utilizing spanning window $(\omega - 1)/2$comments on dual every both word sides. Has been a weight of attention for embedding ithname that estimates the significance of sentence word. The $c(i)$ value, the minimal the word significance. The notation $c(i)$ could be measured with convolution FEATURE matrix $W_L \in \mathbb{R}^{\omega \times d}$ and dependent b_L in the following way as in [2] [62]:|

$$c(i) = \delta\left(e_{L,i} * W_L + b_L\right) \tag{1}$$

$$e_{L,i} = \left(e_{i+\frac{-\omega+1}{2}}, e_{i+\frac{-\omega+3}{2}}, \dots, e_i, \dots, e_{i+\frac{\omega-3}{2}}, e_{i+\frac{\omega-1}{2}}\right)^T \tag{2}$$

The notation * indicates operation of convolution, signifying non-linear function activation that is the function of ReLU in this contribution as in [17]. As per the weight of attention, the matrix of word vector with local weights of attenuation has been expressed in the following

$$\hat{E}_U = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_m) \in R^{\omega \times d} \tag{3}$$

where word vector $\hat{e}_i$ of the i-th word has been measured:

$$\hat{e}_1 = c(i)e_i \tag{4}$$

In the design space, conceivable angle words can be value, class, shading, surface, texture, shape, part, style, and so on, often referenced in client Reviews. As a rule, various clients have various inclinations over style, and a particular client's consideration may change with a focused design. For example, a client may zero in on the quality and t, however, couldn't care less the cost while choosing a suit. A similar client might be more worried about the style and shading than the material when buying a dance skirt. For a particular thing, various clients may buy it with different expectations.

As overall words share similar d measurements across the k perspectives, we utilize the neighbourhood viewpoint exact word change grid W_a^l∈ R^(ω×k)to change the word vector portrayal. W_a^l is instated arbitrarily by a uniform dispersion U(- 0.01,0.01). Hence, we can disengage the neighbourhood word angle portrayal P_(i,a)^l∈R^kfrom e^_ias:

$$P_{i,a}^{l} = \hat{e}_i W_a^{l} \tag{5}$$

Usually, the explicit angle semantics of single word keep an eye on various express polarities in the style area. Along these lines, we should catch this sort of words in the perspective portrayal. For instance, many 'extraordinary' comments in an Review could prompt inverse feelings in such expressions as "an incredible cost" and "ts incredible". Another word, 'high,' conveys various feelings regarding "this clothingwere made with excellent materials, and I would enthusiastically suggest" and "the cost is excessively high."

Thinking about the significance of learning each word in a survey, the nearby viewpoint client portrayal P_a^l∈ R^k can be determined by a consideration component dependent on the accompanying weighted total:

$$P_a^{l} = \sum_i attn_{i,a}^{l} P_{i,a}^{l} \tag{6}$$

$$attn_{i,a}^{l} = softmax(P_{i,a}^{l})^T v_a^{l} \tag{7}$$

Where (the equivalent beneath) is characterized as the I-thneighbourhood consideration vector (for example a likelihood dispersion) over the survey words for a client you concerning viewpoint a. is the nearby angle inserting framework, and it is introduced arbitrarily by a uniform conveyance.

B. GLOBAL PARAMETER EXTRACTION OF FEATURE MODULE

Notwithstanding the neighbourhood perspective element extraction module, an equal module is utilized to separate many worldwide angle client/thing highlights. Motivated by the contribution in [25] & [30], we utilized a bi-LSTM [31] for show worldwide long haul reliance & to discover worldwide angle includes by using both past and future settings and by handling the arrangement in both forward and reverse ways. At each time step t, the yield vectors of 2 headings are linked. Moreover,a worldwide viewpoint highlight extraction module is portrayed in Figure 3. Initially, worldwide long haul reliance of text arrangement data has acquired dependent on the implanted setting. Besides, viewpoint semantic implications of the sentence have been communicated by a global perspective shared change framework. At long last, worldwide angle portrayals are disengaged by a consideration system.

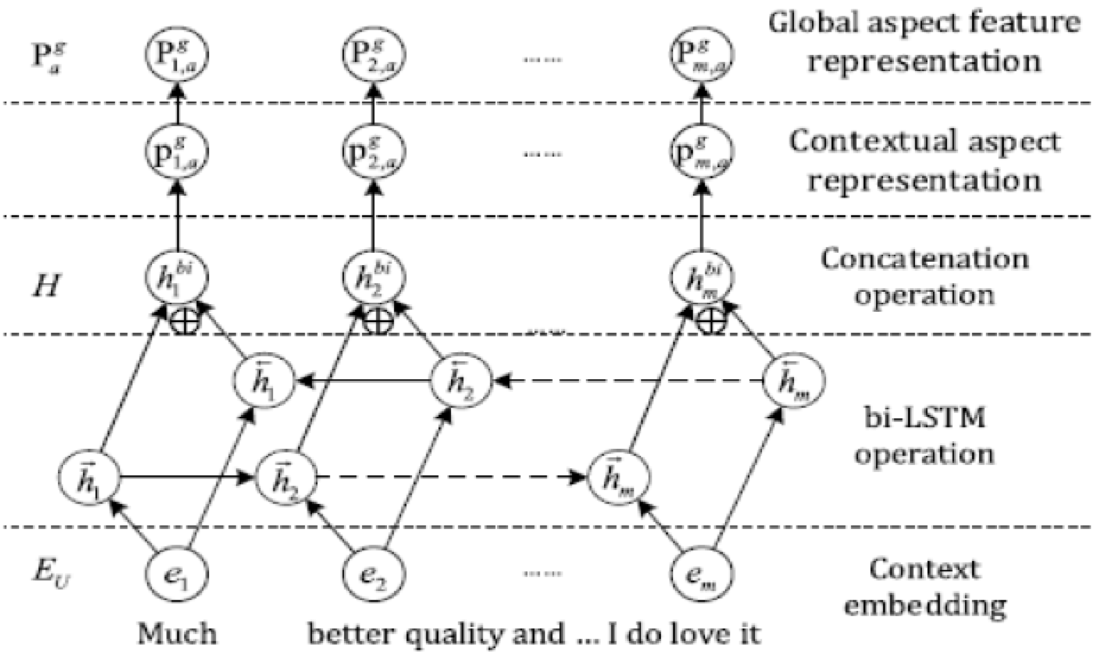


FIGURE 3. Global parameter feature mining module

Let the notation d_h be single direction hidden size LSTM. Also, LSTM hidden state $h_t \in \mathbb{R}^{d_h}$ at the t has been upgraded as:

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

$$o_t = \sigma\,(W_o . E + b_o) \quad (9)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c . E + b_c) \quad (10)$$

$$f_t = \sigma\left(W_f . E + b_f\right) \quad (11)$$

$$i_t = \sigma\,(W_i . E + b_i) \quad (12)$$

$$E = \begin{bmatrix} h_{t-1} \\ e_t \end{bmatrix} \quad (13)$$

Where the notation W_i, W_f, W_o have weighted matrices b_i, b_f, b_o , b_c were dependent as learned at the time of parameterizing and training input gatei_t , ot output-gate, $c_t$state of cell and f_t forget-gate in respective order; the notation h_t has LSTM hidden state; the notation σ has been sigmoid-function; the notation $\odot$ represents for multiplication of element-wise; and the notation e_t has been LSTM cell input unit deliberating the word vectors embedding.

Later, we feed input-embedding sequencing text $E_U = e_1, e_2, \ldots, e_m$ towards LSTM in above-direction and attain the state of forwardinghidden $\vec{h}_t$. One needs to upgrade a backward state of hidden $\overleftarrow{h}_t$ by feeding sequences into LSTM in a backward direction. Also, states of hidden of 2 orders have been aggregated in the following:

$$h_t^{bi} = \vec{h}_t \oplus \overleftarrow{h}_t \quad (14)$$

The notation / indicates overall long-term reliance at the t since it comprises text from two directions. General states of hidden have been gathered into a matrix that stated as:

$$H = [h_1^{bi}, h_2^{bi}, \dots, h_m^{bi}] \quad (15)$$

Where the representation $H \in \mathbb{R}^{m \times 2d_h}$ and every H row depicts the overall long-term reliance at the resulting place of an internal sequence of text

Later, we extract global contextual FEATURE depiction $p_{i,a}^{g}$ from h_i^{bi} by expressing

$$P_{i,a}^{g} = [h_i^{bi}, W_a^{g} \in R^k] \quad (16)$$

Due to the significance of every word learning in reviews differs, the overall user FEATURE depiction P_a^g could be derived by attenuation technique

$$P_a^g = [P_{1,a}^g, P_{2,a}^g, \dots, P_{m,a}^g] \in R^{m \times k} \quad (17)$$

$$P_{i,a}^g = attn_{i,a}^g P_{i,a}^g \quad (18)$$

$$attn_{i,a}^g = softmax\left(\left(P_{i,a}^g\right)^T V_a^g\right) \quad (19)$$

C. MUTUAL OPERATION MODULE

To achieve this approach generalization ability, one needs to aggregate FEATUREs of local and global user parameters by using eq 20 and aggregate it by utilizing equation 21 by an attention technique in 22.

$$\bar{P}_{i,a} = \left[\left(P_a^l - P_{i,a}^g\right) \odot \left(P_a^l - P_{i,a}^g\right)\right] \in R^k \quad (20)$$

$$\hat{P}_{u,a} = \sum_i attn_{i,a} \bar{P}_{i,a} \in R^k \quad (21)$$

$$attn_{i,a} = softmax\left(\bar{P}_{i,a}\right) \quad (22)$$

$$\gamma_{u,a} = \sigma\left(\hat{P}_{u,a} W_1 + b^1\right) W_2 + b^2 \quad (23)$$

For enhancing the performance of generalization, one needs to modify the dropout strategy that has been utilized extensively in contemporary neural approaches for the suggestion as in [2] [4] [16]

D. ESTIMATION RATING OPTIMIZATION

There are two parallel channels used for learning depictions of user features parameters and features item parameter $\gamma_{i,a}$ and $\gamma_{u,a}$ that could be integrated for forming total rating $\hat{r}$, in the following:

$$\hat{r} = \sum_a \left(\gamma_{u,a}\right)^T \gamma_{i,a} + b_u + b_i + b_0 \quad (24)$$

Where b_u, b_i, and b_0 were the item ofuser, & global biases [4], in respective order.

The above prediction could be deliberated in the form of the regression issue. Overall FEATUREs have been combined trained by back-propagation strategy [62], where MSE (mean-square-error) has been utilized in the form of loss-function. For learning this approach FEATUREs, the function of objective, J could be stated as

$$J = \sum (r - \hat{r})^2 + \lambda_{\ominus} \|\ominus\|^2 \quad (25)$$

4. EXPERIMENTS

In this part, exhaustive investigations on two open survey datasets are introduced to assess the presence of FFRAM. The data about the datasets, the standard strategies, the examination arrangement, and the outcomes are explained.

A DATASETS

The two datasets review utilized in the simulation isjewellery, shoes, and clothing dataset from 5-core Amazon [63] and dataset from 1 US online-retailer as in D2. Two datasets have been filtered to assure that every item or user review has a minimal 1 rating. Moreover, datasets' fundamental characteristics have been exhibited in the table below, where #user, # item, and # rating were the number of things, ratings and users in every dataset, $\text{Density} = \#\text{Rating} / (\#\text{User} \times \#\text{Item})$, & sparsity = 1-Density. It might be observed that D2 is ten times greater than users D1. Both have nearly the same amount of items. The D2's # rating has been two times of D2 & D1 density has been approximately D1 of 1/5.

TABLE 1. Statistics of datasets used in this paper.

Dataset	#User	#Item	#Rating	Density (%)	Sparsity (%)
D1	48.235	21000	215831	0.058	99.258
D2	451089	32518	254871	0.0025	99.584

B. BASELINES

For examining the projected approach performance, FFRAM approach has been utilized the outcomes of below rating estimation approaches in the form of baselines.

1) DeepCoNN [17]: DeepCNN has been dependent on two parallel CNNs for learning vectors of the latent feature of user & their item from the point of view& the FM for predict predictions.

2) D-Attn [2]: This approach includes global and local attention depending on modules for choosing globally as well as locally informative reviews words and for attaining the user latent features interpretability of item and user reviews.

3) NARRE [16]: The NARRE approach exploits two attention techniques and parallel CNNs for learning user's latent-features and reviews of items for completing the estimation of rating T.

4) ANR [4]: ANR has been dependent on neural co-attention and attention ideas by incorporating FEATURE from learning device and the predictor of the significance of factor.

5) DAML [62]: This approach uses mutual and local attention from CNN's techniques and MLP to attain a predictive user rating.

C. EXPERIMENT SETUP

In the simulation with FFRAM, one needs to set latent vector feature set of the user along with reviews of the user at the value 300, a window size of sliding at five value, 0.5 could be the rate of dropout, length of training batch could be 64, amount of parameters at the 5 for D2 and D1. Moreover, for D1, the learning rate has been set 0.0001, & size of LSTM at the 5. Aimed at D2, quality of learning has been developed over 0.0008 and bi-LSTM hidden size as 7

The MAE and MSE have been utilized as assessment metrics. The simulation has been repeated for five times for attaining constant outcomes when verifying MSE as minimal and average MAE and MSE of 5 test have been computed. Our simulation has performed with python 3.6 as well as 0.41 PyTorch on NVIDIA

D. COMPARISON OF PERFORMANCE

The FFRAM performance on two specified datasets have been contrasted, and five cutting proposal approaches. Tab 2 exhibits trail outcomes from where the optimal result has been attributed by robust and next optimal by highlighted. 1% means the usual distinction in MAE or MSE among FFRAM and pattern technique that measures the presentation enhancement by FFRAM.

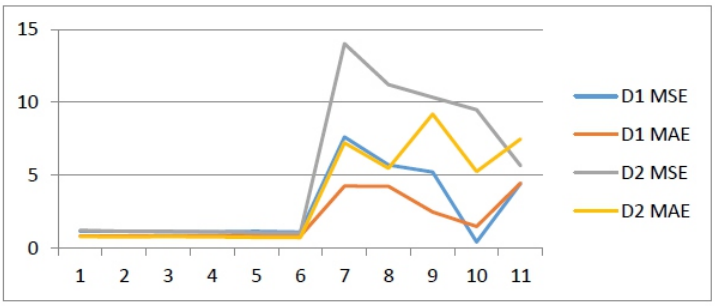
TABLE 2. MSE and MAE performance on two datasets.

Methods		D1		D2	
		MSE	MAE	MSE	MAE
Error	(a)	1.1478	0.8415	1.2148	0.7854
	(b)	1.1584	0.8258	1.1458	0.7648
	(c)	1.1125	0.8369	1.1487	0.7928
	(d)	1.1118	0.8248	1.1249	0.7647
	(e)	1.1589	0.8258	1.0478	0.734
	(f)	1.0984	0.8248	1.0248	0.7248
Δ%	(a)	7.61	4.25	14.00	7.21
	(b)	5.70	4.24	11.21	5.47
	(c)	5.21	2.48	10.34	9.18
	(d)	0.43	1.48	9.48	5.24
	(e)	4.42	4.47	5.64	7.47

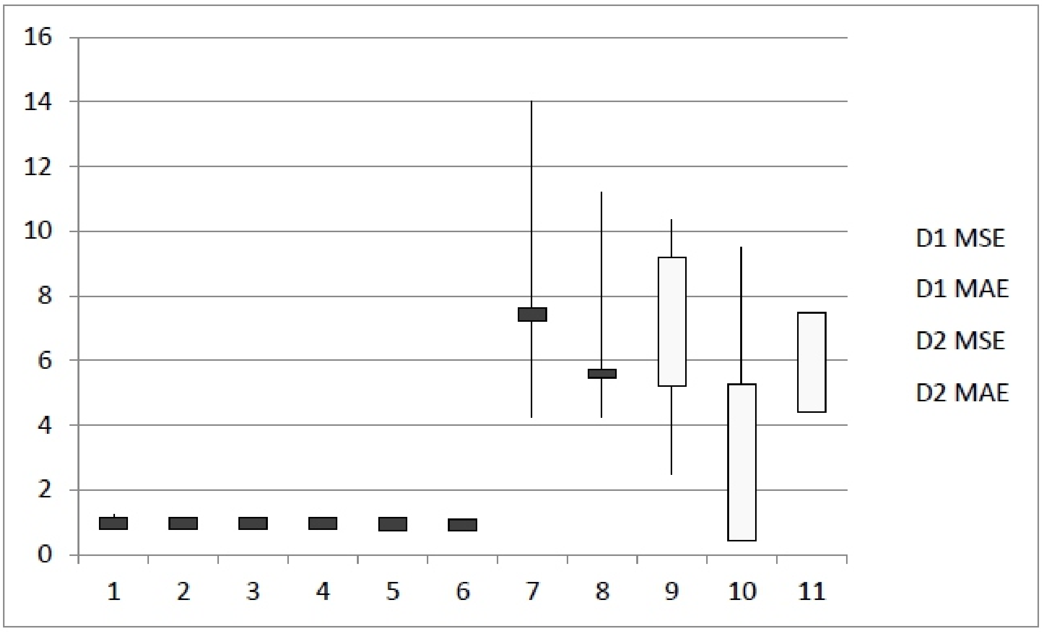
The outcomes exhibit that FFRAM attained the optimal MAE and MSE scores and performed well for both the datasets. Also, the 2nd optimal performer has been ANR; on D1 the FFRAM enhanced MSE with 0.43% and MAE through 1.51 in respective order. For the D2, the next optimal performer has been DAML when MSE has been regarded. And when MAE has been concerned. However, FFRAM enhanced MAE and MSE from 2nd optimal performed by 5.14 and 5.74% in respective order. The simulation exhibited that model based on the FEATURE with attention technique among item and user reviews might offer more precise rating estimations than other approaches on two datasets.

5. MODELANALYSIS A. INFLUENCES OF HIDDEN SIZE & NUMBER OF FEATURES

They chooseDH, size of hidden, and amount of parameters K possessing straight influences on rating predictions. Moreover, in our simulation, FFRAM hidden size has been selected from the [1, 3, 5, 7, 9] & amount of FEATURE differs from 1-8. Fig 4 exhibits the impacts of k andDH on estimated ratings for reviews in 2 datasets. Furthermore, graphs show optimum performances happened on D1.



(a)



(b)

Figure 4. Influences of hidden size and Features. (a) Amazon dataset (D1). (b) 2nd online retailer's dataset (D2).

B. MODEL INTERPRETABILITY

Identical to which has identified in contribution [4] [6], five fashion parameters have been frequently perceived amid reviews in 2 datasets $\mathcal{A} = \{\text{size/fit, color, fabric/texture, price, style}\}$. Moreover, the distribution background of every embedding e_i word has been determined in the form of $b_e = \sum_{a\in\mathcal{A}} \psi_a / |\mathcal{A}|$, where $\psi_a = \sum_{u\in U} \psi_{u,a} / |\mathcal{U}| + \sum_{i\in I} \psi_{i,a} / |\mathcal{I}|$. ψ_a indicates to the significance of the name for the parameter a, as well as $\psi_{u,a} = \sum_i attn_{i,a}$ has been a sign of every e_i words through in respect to consumer $attn_{i,a}$ and parameter $u \in \mathcal{U}$ over V vocabulary. Hence, we might depict parameter by utilizing their top

TABLE 3. Top 8 words of each FEATURE in D1.

Size/fit	Colour	Fabric/texture	Price	Style
Length	colours	material	price	style
Medium	black	shirt	deal	comfort
Usually	white	soft	cost	classic
Sizes	design	fabric	value	particular
normally	blue	cotton	paid	fashion
shape	red	lace	reasonable	unique
sizing	brown	sweater	bargain	statement
fitted	dark	smooth	unbeatable	styling

TABLE 4. Top 8 words of each FEATURE on D2.

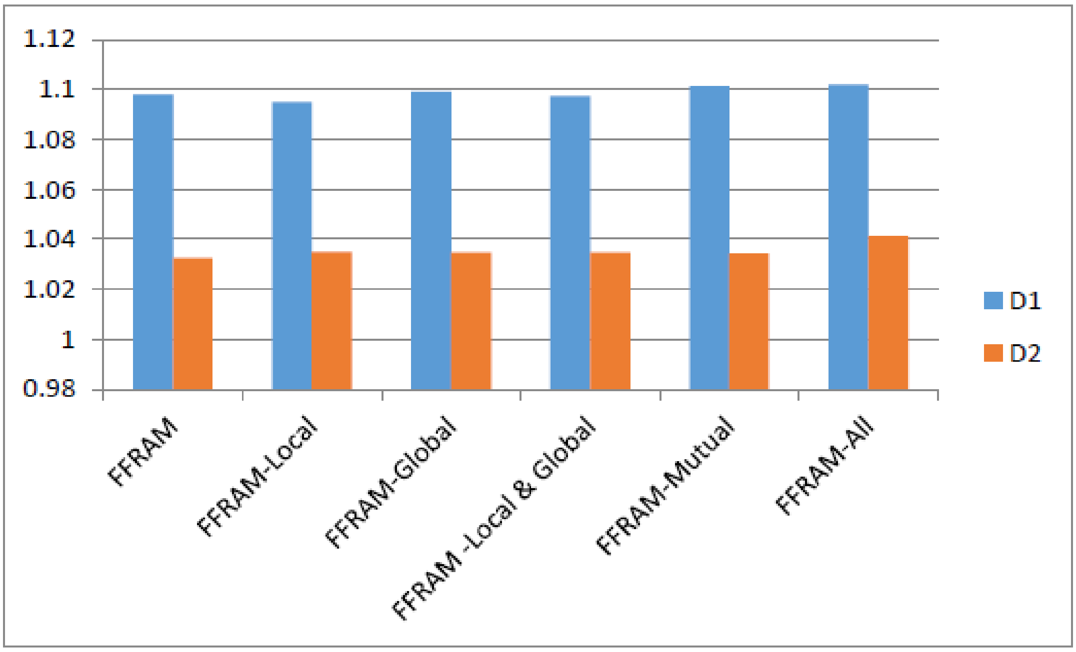
Size/fit	Colour	Fabric/texture	Price	Style
fit	colour	soft	price	style
size	colours	smooth	quality	comfort
fits	shade	material	value	classic

usually medium length shape sizes	dark black blush pink white	shirt fabric thin sheer cotton	deal cost reasonable paid prices	unique styles particular styling fashion

TABLE 5. Influence of attention layers in FFRAM on MSE.

Dataset	FFRAM	FFRAM-Local	FFRAM-Global	FFRAM -Local & Global	FFRAM-Mutual	FFRAM-All
D1	1.0978	1.09478	1.0991	1.0972	1.1012	1.1019
D2	1.0324	1.03489	1.0348	1.0348	1.0344	1.0414

Every FEATURE. Moreover, these words in 5 parts reflect the associations amid users, rating & reviews.



C. ATTENTION LAYER IMPACT

We examined the influences of local, mutual and global layer attention. Tab 5 offers performance comparison while attention-layers in the FFRAM-varied other FEATUREs settings rest to be similar. The FFRAM global and local deprived of international or local layer attention. The global and regional FFRAM deprived of attention global and local layers. Also, FFRAM-all and FFRAM mutual have FFRAM deprived of joint layer attention or overall layers of attention.

FFRAM can obtain the best MSE on the two datasets when all the attention layers were included, and the worst MSE when all the attention layers were removed on both datasets. In the other four scenarios, attention mechanisms performed differently on D1 and D2. Removing the mutual attention mechanism had the most significant impact on MSE, indicating that the attention layers, especially the joint attention layer, can improve the recommendation performance. This is because the mutual attention layer can combine different word polarities in a sentence and the whole sentence's FEATURE semantic meanings.

6. CONCLUSION

This paper introduced a perspective-based style proposal model with consideration instrument (FFRAM) to foresee clients' evaluations dependent on clients' surveys of bought design items. This model utilized two similar ways to remove inert viewpoint highlights about clients and things independently and a shared activity module to consolidate the two ways toward the end for anticipating clients' evaluations. On every way, there was a convolutional neural organization (CNN) and a long transient memory (LSTM) organization, both having a consideration component, to catch neighbourhood viewpoint highlights and worldwide perspective highlights at the same time. The typical tasks consolidating nearby and worldwide perspective highlights in both client and thing Review significantly improved the FFRAM model's speculation. As exhibited in the examination with certifiable client Reviews and evaluations gathered from two famous business sites, FFRAM beat the five best in class recommenders as far as the exactness of foreseeing client appraisals on design items.

REFERENCES

1. *Y. Lu, R. Dong, and B. Smyth, ``Coevolutionary recommendation model: Mutual learning between ratings and reviews,'' in Proc. World Wide Web Conf., 2018, pp. 773-782.*
2. *S. Seo, J. Huang, H. Yang, and Y. Liu, ``Interpretable convolutional neural networks with dual local and global attention for review rating prediction,'' in Proc. 11th ACM Conf. Recommender Syst., Aug. 2017, pp. 297-305.*
3. *R. Catherine and W. Cohen, ``TransNets: Learning to transform for the recommendation,'' in Proc. 11th ACM Conf. Recommender Syst., Aug. 2017, pp. 288-296.*
4. *J. Y. Chin, K. Zhao, S. Joty, and G. Cong, ``ANR: FEATURE-based neural recommender,'' in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2018, pp. 147-156.*
5. *M. Blei, A. Y. Ng, and M. I. Jordan, ``Latent Dirichlet allocation,'' J. Mach. Learn. Res., vol. 3, pp. 993-1022, Mar. 2003.*
6. *J. McAuley and J. Leskovec, ``Hidden factors and hidden topics: understanding rating dimensions with review text,'' in Proc. 7th ACM Conf. Recommender Syst., 2013, pp. 165-172.*
7. *Ling, M. R. Lyu, and I. King, ``Ratings meet reviews, a combined approach to recommend,'' in Proc. 8th ACM Conf. Recommender Syst., 2014, pp. 105-112.*
8. *Y. Bao, H. Fang, and J. Zhang, ``Topicmf: Simultaneously exploiting ratings and reviews for the recommendation,'' in Proc. 28th AAAI Conf. Artif. Intell., vol. 2014, pp. 2-8.*
9. *K. Song, W. Gao, S. Feng, D. Wang, K.-F. Wong, and C. Zhang, ``Recommendation vs sentiment analysis: A text-driven latent factor model for rating prediction with cold-start awareness,'' in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 2744-2750.*
10. *Y. Kim, ``Convolutional neural networks for sentence classification,'' in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp., 1746-1751.*
11. *W. Ouyang, B. Xu, J. Hou, and X. Yuan, ``Fabric defect detection using activation layer embedded convolutional neural network,'' IEEE Access, vol. 7, pp. 70130-70140, 2019.*
12. *S. Hochreiter and J. Schmidhuber, ``Long short-term memory,'' Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.*
13. *T. Mikolov, M. Karaát, L. Burget, J. Ernocký, and S. Khudanpur, ``Recurrent neural network-based language model,'' in Proc. 11th Annu. Conf. Int. Speech Commun. Assoc., 2010, pp., 1045-1048.*
14. *S. Sabour, N. Frosst, and G. E. Hinton, ``Dynamic routing between capsules,'' in Proc. Adv. Neural Inf. Process. Syst., 2017, pp., 3856-3866.*
15. *Li, C. Quan, L. Peng, Y. Qi, Y. Deng, and L.Wu, ``A capsule network for recommendation and explaining what you like and dislike,'' in Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2019, pp. 275-284.*
16. *C. Chen, M. Zhang, Y. Liu, and S. Ma, ``Neural attentional rating regression with review-level explanations,'' in Proc. World Wide Web Conf., 2018, pp. 1583-1592.*
17. *L. Zheng, V. Noroozi, and P. S. Yu, ``Joint deep modelling of users and items using reviews for the recommendation,'' in Proc. 10th ACM Int. Conf. Web Search Data Mining, 2017, pp. 425-434.*
18. *Almahairi, K. Kastner, K. Cho, and A. Courville, ``Learning distributed representations from reviews for collaborative filtering,'' in Proc. 9th ACM Conf. Recommender Syst., 2015, pp. 147-154.*
19. *Q. Cui, S.Wu, Q. Liu,W. Zhong, and L.Wang, ``MV-RNN: A multi-view recurrent neural network for the sequential recommendation,'' IEEE Trans. Knowl. Data Eng., vol. 32, no. 2, pp. 317-331, Feb. 2020.*

20. C. He, Y. Liu, Q. Guo, and C. Miao, ``Multi-scale quasi-RNN for next item recommendation,'' 2019, arXiv:1902.09849. [Online]. Available: http://arxiv.org/abs/1902.09849
21. Q. Cui, S. Wu, Y. Huang, and L. Wang, ``A hierarchical contextual attention-based network for the sequential recommendation,'' Neurocomputing, vol. 358, pp. 141-149, 2019.
22. W.-C. Kang and J. McAuley, ``Self-attentive sequential recommendation,'' in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2018, pp. 197-206.
23. R. Huang, S. McIntyre, M. Song, H. E, and Z. Ou, ``An attention-based recommender system to predict contextual intent based on choice histories across and within sessions,'' Appl. Sci., vol. 8, no. 12, p., 2426, Nov. 2018.
24. Y. Wang, A. Sun, J. Han, Y. Liu, and X. Zhu, ``Sentiment analysis by capsules,'' in Proc. World Wide Web Conf., vol. 2, 2018, pp., 1165-1174.
25. Q. Ma, L. Yu, S. Tian, E. Chen, and W. W. Y. Ng, ``Global-local mutual attention model for text classification,'' IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 27, no. 12, pp., 2127-2139, Dec. 2019.
26. S. Hochreiter, ``The vanishing gradient problem during learning recurrent neural nets and problem solutions,'' Int. J. Uncertainty, Fuzziness Knowl.- Based Syst., vol. 6, no. 2, pp. 107-116, Apr. 1998.
27. R. Pascanu and T. Mikolov, ``On the difficulty of training recurrent neural networks,'' in Proc. 30th Int. Conf. Mach. Learn., 2013, pp., 1310-1318.
28. T. Donkers, B. Loepp, and J. Ziegler, ``Sequential user-based recurrent neural network recommendations,'' in Proc. 11th ACM Conf. Recommender Syst., Aug. 2017, pp. 152-160.
29. S. Heinz, C. Bracher, and R. Vollgraf, ``An LSTM-Based dynamic customer model for fashion recommendation,'' in Proc. CEUR Workshop Proc., vol. 1922, 2017, pp. 45-49.
30. J. Zeng, X. Ma, and K. Zhou, ``Enhancing attention-based LSTM with position context for FEATURE-level sentiment classification,'' IEEE Access, vol. 7, pp. 20462-20471, 2019.
31. M. Schuster and K. K. Paliwal, ``Bidirectional recurrent neural networks,'' IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673-2681, 1997.
32. Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, ``Explicit factor models for an explainable recommendation based on phrase-level sentiment analysis,'' in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2014, pp. 83-92.
33. N. Wang, H. Wang, Y. Jia, and Y. Yin, ``Explainable recommendation via multi-task learning in opinionated text data,'' in Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jun. 2018, pp. 165-174.
34. X. He, T. Chen, M.-Y. Kan, and X. Chen, ``TriRank: Review-aware explainable recommendation by modelling FEATUREs,'' in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. 2015, pp., 1661-1670.
35. K. Bauman, B. Liu, and A. Tuzhilin, ``FEATURE based recommendations: Recommending items with the most valuable FEATUREs based on user reviews,'' in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2017, pp. 717-725.
36. X. Guan, Z. Cheng, X. He, Y. Zhang, Z. Zhu, Q. Peng, and T.-S. Chua, ``Attentive FEATURE modelling for the review-aware recommendation,'' ACM Trans. Inf. Syst., vol. 37, no. 3, pp. 1-27, Jul. 2019.
37. Y. Wu and M. Ester, ``FLAME: A probabilistic model combining FEATURE-based opinion mining and collaborative filtering,'' in Proc. 8th ACM Int. Conf. Web Search Data Mining, 2015, pp. 199-208.
38. J. Zhang, D. Chen, and M. Lu, ``Combining sentiment analysis with a fuzzy Kano model for product FEATURE preference recommendation,'' IEEE Access, vol. 6, pp. 59163-59172, 2018.
39. Z. Cheng, Y. Ding, L. Zhu, and M. Kankanhalli, ``FEATURE-aware latent factor model: Rating prediction with ratings and reviews,'' in Proc. World Wide Web Conf. World Wide Web, 2018, pp. 639-648.
40. Sutskever, O.Vinyals, and Q.V. Le, ``Sequence to sequence learning with neural networks,'' in Proc. Adv. Neural Inf. Process. Syst., vol. 4, Sep. 2014, pp., 3104-3112.
41. Xu, ``Show, attend and tell: Neural image caption generation with visual attention,'' Comput. Sci., vol. 2015, pp. 20482057, Feb. 2015.
42. X.Wang, L. Yu, K. Ren, G. Tao,W. Zhang, Y. Yu, and J.Wang, ``Dynamic attention deep model for article recommendation by learning human Editors' demonstration,'' in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2017; pp., 2051-2059.
43. Li, M. R. Min,Y. Ge, and A. Kadav, ``Acontext-aware attention network for interactive question answering,'' in Proc. 23rd ACMSIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2017, pp. 927-935.
44. P. Ren, Z. Chen, Z. Ren, F. Wei, J. Ma, and M. de Rijke, ``Leveraging contextual sentence relations for extractive summarization using a neural attention model,'' in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2017, pp. 95-104.
45. M. Xu Chen, O. Firat, A. Bapna, M. Johnson, W. Macherey, G. Foster, L. Jones, N. Parmar, M. Schuster, Z. Chen, Y. Wu, and M. Hughes, ``The best of both worlds: Combining recent advances in neural machine translation,'' 2018, arXiv:1804.09849. [Online]. Available: http://arxiv.org/abs/1804.09849

46. Serrá, D. Suris, M. Miron, and A. Karatzoglou, ``Overcoming catastrophic forgetting with hard attention to the task,'' in Proc. 35th Int. Conf. Mach. Learn., 2018, pp., 4548-4557.
47. Y. Tay, A. T. Luu, and S. C. Hui, ``Multi-pointer co-attention networks for the recommendation,'' in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2018, pp., 2309-2318.
48. J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T.-S. Chua, ``Attentive, collaborative altering: Multimedia recommendation with Item- and component-level attention,'' in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2017, pp. 335-344.
49. X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, ``KGAT: Knowledge graph attention network for the recommendation,'' in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2019, pp. 950-958.
50. S. Shankar and S. Sarawagi, ``Posterior attention models for the sequence to sequence learning,'' in Proc. 7th Int. Conf. Learn. Represent.,2019, pp. 1-11.
51. Q. Zhang, Y. Gong, J. Wu, H. Huang, and X. Huang, ``Retweet prediction with the attention-based deep neural network,'' in Proc. 25th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2016, pp. 75-84.
52. S. Kang, T. Phan, M. Bolas, and D. M. Krum, ``Spatio-temporal wardrobe generation of actors' clothing in video content,'' in Proc. Int. Conf. Hum. -Comput. Interact. 2016, pp. 448-459.
53. He, S. Zhang, and Q. Meng, ``Learning to style-aware Bayesian personalized ranking for the visual recommendation,'' IEEE Access, vol. 7, pp. 1419814205, 2019.
54. Tuinhof, C. Pirker, and M. Haltmeier, ``Image-based fashion product recommendation with deep learning,'' in Proc. Int. Conf. Mach. Learn., Optim., Data Sci., 2018, pp. 472-481.
55. G.-L. Sun, Z.-Q. Cheng, X. Wu, and Q. Peng, ``Personalized clothing recommendation combining user social circle and fashion style consistency,'' Multimedia Tools Appl., vol. 77, no. 14, pp. 1773117754, Jul. 2018.
56. Y. Zhang, X. Liu, Y. Shi, Y. Guo, C. Xu, E. Zhang, J. Tang, and Z. Fang, ``Fashion evaluation method for clothing recommendation based on weak appearance feature,'' Sci. Program., vol. 2017, pp. 1-12, Oct. 2017.
57. T. Nguyen, T. Almenningen, M. Having, A.Kofod-Petersen, H. Langseth, and H. Ramampiaro, ``Learning to rank for personalized fashion recommender systems via implicit feedback,'' in Proc. Mining Intell. Knowl. Explore., 2014, pp. 51-61.
58. Z. Cheng, X. Chang, L. Zhu, R. C. Kanjirathinkal, and M. Kankanhalli, ``MMALFM: Explainable recommendation by leveraging reviews and images,'' ACM Trans. Inf. Syst., vol. 37, no. 2, pp. 128, Mar. 2019.
59. Q. Liu, S. Wu, and L. Wang, ``DeepStyle: Learning user preferences for the visual recommendation,'' in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2017, pp. 841-844.
60. W. Yu, H. Zhang, X. He, X. Chen, L. Xiong, and Z. Qin, ``Aesthetic-based clothing recommendation,'' in Proc. World Wide Web, 2018, pp. 649-658.
61. Google. word2vec. Accessed: Jul. 29, 2013. [Online]. Available: https://code.google.com/archive/p/word2vec/
62. Liu, J. Li, B. Du, J. Chang, and R. Gao, ``DAML: Dual attention mutual learning between ratings and reviews for an item recommendation,'' in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2019, pp. 344-352.
63. Amazon. Amazon Product Data. Accessed: Feb. 14, 2020. [Online]. Available: http://jmcauley.ucsd.edu/data/amazon

ISSN 2210-142X

Discovering hidden topics in Moroccan News published online using BERT and Neural Topic Model

Nassera Habbat[1], Houda Anoun[2] and Larbi Hassouni[3]
[1-2-3] RITM Laboratory, CED ENSEM Ecole Superieure de Technologie Hassan II University
E-mail address: nassera.habbat@gmail.com, houda.anoun@gmail.com, lhassouni@hotmail.com

ABSTRACT

Topic modeling algorithms can better understand data by extracting meaningful words from text collection, but the results are often inconsistent, and consequently difficult to interpret. Enrich the model with more contextual knowledge can improve coherence. Recently, neural topic models have emerged, and the development of neural models, in general, was pushed by BERT-based representations. We propose in this paper, a model named AraBERTopic to extract news from Facebook pages. Our model combines the Pre-training BERT transformer model for the Arabic language (AraBERT) and neural topic model ProdLDA. Thus, compared with the standard LDA, pre-trained BERT sentence embeddings produce more meaningful and coherent topics using different embedding models. Results show that our AraBERTopic model gives 0.579 in topic coherence.

Keywords: Neural topic model, ProdLDA, AraBERT, topic coherence.

1. INTRODUCTION

Nowadays, because of the exponential development of the Internet, a huge quantity of documents, such as online news are produced every day, especially in social media like Facebook which is one of the most important social networks in Africa. Concerning Morocco, there were 22 010 000 Facebook users in January 2021 [1]. Mining the knowledge and topics from social media posts have attracted a lot of attention these last years. To discover hidden topical patterns which are present in large collections of texts, many unsupervised techniques are usually used to generate probabilistic topic models. Among these models, we find Non-negative Matrix Factorization (NMF), Latent Semantic Indexing (LSI), and Latent Dirichlet Allocation (LDA) [2].

However, those probabilistic models do not take advantage of language model pre-training benefits. Many extensions were proposed to integrate different types of information and add contextual knowledge to the topic models. The most prominent architecture in this category is Bidirectional Encoder Representations from Transformers (BERT) which allows us to extract representations from pre-trained documents to easily reach state-of-the-art performance through numerous tasks [3].

Recently, some neural topic models were explored and have shown promising results. For instance, ProdLDA, which is a developed version of LDA based on deep learning, is an expert products instead of mixture model in LDA to yield much more interpretable topics [4].

The main contribution of this paper is to present a proposed model (AraBERTopic) to extract topics from Arabic news published on Facebook pages using AraBERT as language model pre-training and ProdLDA as a neural topic model. To prove the high performance of our model; we compared, on the one hand, its feature extraction phase with different embedding models (Glove, Doc2Vec, and Asafaya as a bert-base-arabic model [5]), and we compared, on the other hand, its topic model phase with standard

LDA. The results demonstrate that our proposed model is superior to other models in terms of Normalized topic coherence, Pointwise Mutual Information (NPMI), and perplexity metrics.

The rest of this paper is organized as follows: Section 2 provides a brief review of literature. The proposed model is described in 3rd Section. In section 4, we present the results and discusses. Finally, the paper is summarized and the future work has prospected.

2. RELATED WORKS

We introduce in this part some topic extraction methods, including traditional feature extraction methods and deep learning methods.

Probabilistic topic models are much used in natural language processing (NLP), and among the most commonly used methods, we find LDA. In [6]; the author reviews the academic papers on LDA topic modeling published from 2003 to 2016.

To imitate the statistical process of LDA, the authors in [7] investigate the possibility of using deep neural network to model the statistical process to minimize the computational time in LDA; Therefore, they proposed two deep neural network variants: two and three Neural Network (NN) DeepLDA, in their experiments they used Reuters-21578 as a dataset, and some standard libraries in Python like genism, NLTK and Keras and to record the accuracy of the models, a Support Vector Classifier (SVC) was used. Their results showed that 3NN DeepLDA outperforms 2NN DeepLDA and LDA.

In recent years, deep learning has become a powerful machine learning technology, which allows learning multi-level representation and several methods exist that are particularly adapted for learning meaningful hidden representations. Among those models, we find the Variational Autoencoder (VAE) which is a deep generative model. In [4] the authors present the first efficient autoencoding variational Bayes (AEVB) which is an inference method based on latent Dirichlet allocation (LDA)-(AVITM), in this paper, they proposed a novel topic model named ProdLDA which replace the hybrid model in LDA with expert products, After applying their model on 20 Newsgroup dataset they obtained that AVITM outperforms baseline methods in term of accuracy and reasoning time, and the topics given by ProdLDA are more explanatory, Similarly, the authors in [8] presented Neural Variational Correlated Topic Model consisting of two main parts; the 1st one is the inference network with Centralized Transformation Flow and the 2nd one is the multinomial softmax generative model. To evaluate their model they used NPMI topic coherence. Their results showed that the model enhances the performance of topic modeling and can effectively capture topic correlation.

Only some recent studies have used semantic embeddings like Bidirectional Encoder Representations from Transformers (BERT) and ELMO (Embeddings from Language Models) in topic analysis. In [9] the authors proposed Variational Auto-Encoder Topic Model (VAETM) which combines entity vector representation and word vector representation. The model uses large-scale external corpus and manually edited large-scale knowledge graph to learn the embedding representation of each word and entity, then, those embedding representations are integrated into the VAE framework to deduct the hidden representation of topic distributions; To prove the performance of their model they compared it with various of baseline algorithms (LDA, Sparse Additive Generative Model (SAGE) [10] and SCHOLAR [11]), and 20Newsgroups. IMDB and Chinese Standard Literature (96,000 national and industry standards of China) using as datasets; based on perplexity, NPMI, and accuracy measures; they showed that the model better mine the hidden semantic of short texts and improve topic modeling.

Many researchers opt for BERT [3] as contextualized word representations because it progresses the state of the art for different NLP tasks by pushing the MultiNLI accuracy to 86.7%, score to 80.5% for the General Language Understanding Evaluation (GLUE), the Stanford Question Answering Dataset (SQuAD v1.1) question answering Test F1-score to 93.2 and Test F1-score to 83.1 for SQuAD v2.0, compared to Glove, ELMOs, and OpenAI GPT; Among NLP tasks we found Topic modeling. The authors in [12] include in a neural topic model, the contextualized BERT embeddings to get more consistent topics compared to Neural-ProdLDA, NVDM, and LDA.

Moreover, BERT creates word embeddings in multiple languages, in [13] the authors used LDA topic model and multilingual pre-trained BERT embeddings to analyze the evolution of topics in Chinese, English, and multilingual in scientific publications using Google-pre-trained BERT models: "bert-base-chinese" for Chinese, "bert-base-uncased" for English and ¨bert-base-multilingual-uncased" for multilingual text. The results showed that the model can well analyze the scientific evolution of similar relationships between monolingual and multilingual disciplines. In most cases, 80% of the relationships are related to the key topics of each language.

Compared with English, Arabic is a language with rich forms, less syntactic exploration, and fewer resources. The pre-trained AraBERT model [14] is very effective in language understanding compared with multilingual BERT. It achieves the most advanced performance in most Arabic NLP tasks.

3. PROPOSED APPPROCH

We will describe in this part, our model (AraBERTopic), the global architecture is exposed in Fig. 1. There are three principal components:

(1) Data Acquisition: This component aims to collect data from Facebook pages using web scraping;

(2) Pre-training model using AraBERT: Extract the features.

(3) Extraction of topics using Neural Prod-LDA

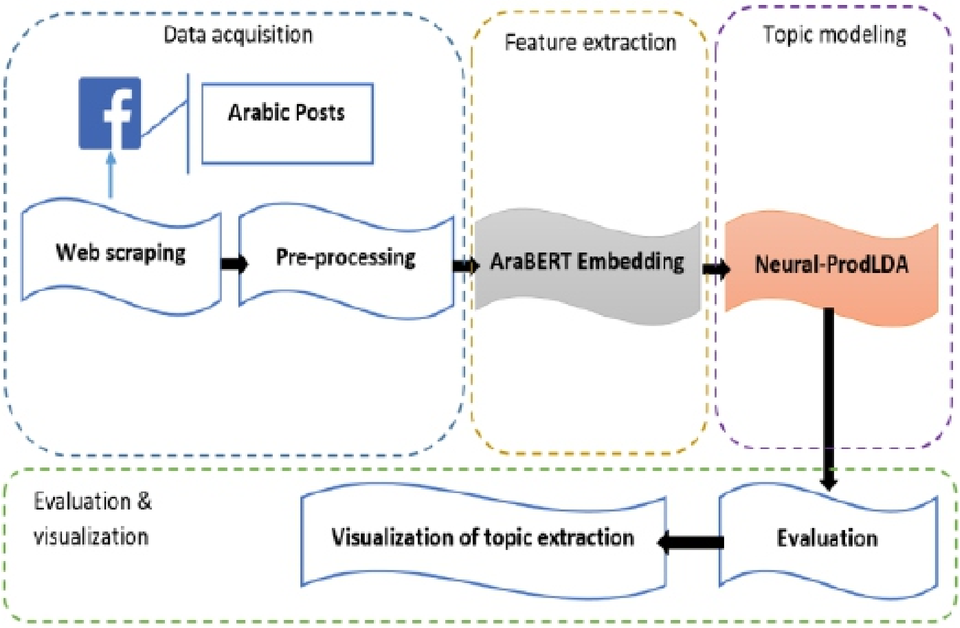


Figure 1. The global architecture of AraBERTopic model.

A. Word embeddings

AraBERT [14] is a pre-training BERT transformer model for the Arabic language. It uses Transformer to learn the context between words (or sub-words) in texts. The Transformer encoder is considered to be bidirectional, which means that it reads the whole word sequence at one time, instead of reading the text input-oriented model in order (from right to left or from left to right).

The input is a token sequence, which is first embedded into a vector and then processed in the neural network to obtain a vector sequence as output. As shown in Figure 2, [CLS] is a special symbol added before each sample input and [SEP] is a special separator mark.

Generally, the embedding models include the following three layers:

- The token embeddings layer changes each word tag into a 768-dimensional vector representation.
- The segment embeddings layer has two representations Vector: the 1st vector (index 0) is attributed to all tokens of input 1, and the last vector (index 1) is attributed to all tokens of input 2.
- The Position embeddings layer: AraBERT is designed to process up to 512 input sequences. Therefore, AraBERT must learn a vector representation of each position. This means that the position integration layer is a size look-up table (512, 768) where :
- The 1st row is the representation of the vector of each word in position 1.
- The 2nd row is the representation of the vector of each word in position 2, etc.

These representations are added in a single representation by the elements in the generated form (1, n, 768). The following Figure shows the input representation passed to the AraBERT encoder layer.

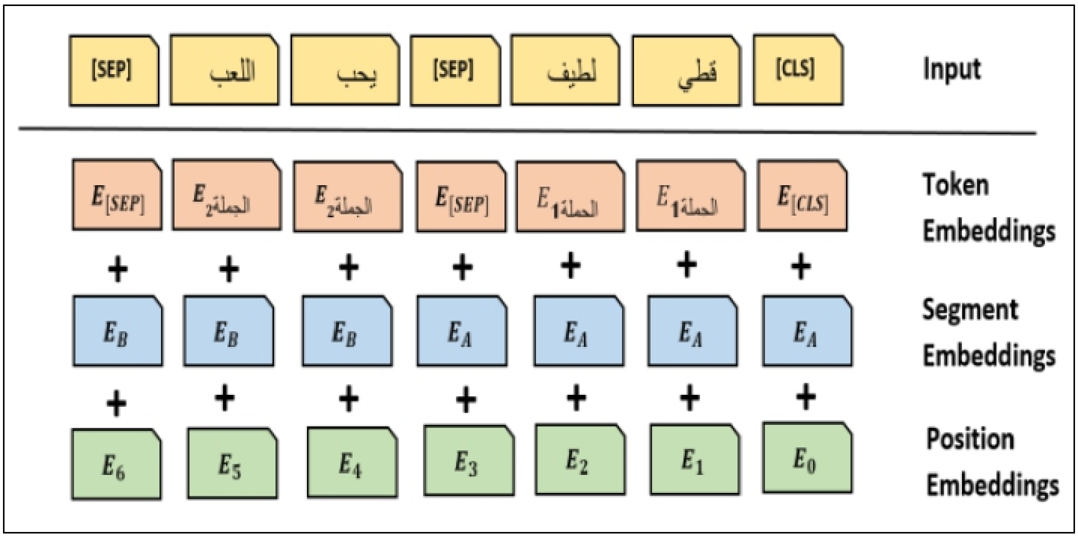


Figure 2. AraBERT input representation.

The learning of AraBERT takes place in two phases:

The 1st one is the pre-training, it is done only once, it allows the creation of a neural network that has a certain general understanding of the language.

Then, the 2nd phase is called the fine-tuning phase, which allows the network to be trained on a specific task like classification, question answering, and Topic modeling as shown in figure 3.

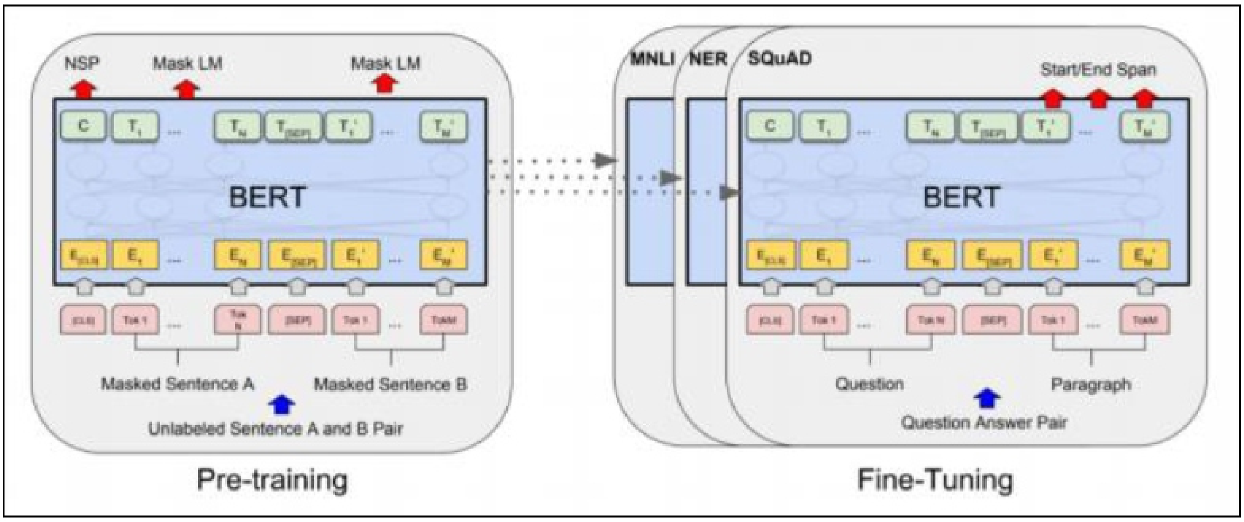


Figure 3. General pre-training and fine-tuning processes for BERT [3].

Concerning Pre-training Dataset, it was scraped from Arabic news websites for articles. In addition there are two major publicly accessible Arabic corpora:

- ✓ Open Source International Arabic News (OSLAN) Corpus [14] composed of 3.5 million articles (about 1 billion tokens) from 31 news sources in 24 Arab countries,
- ✓ An Arabic corpus of 1.5 billion word [13], which is a contemporary corpus composed of more than 5 million articles from ten major information sources in 8 countries,

After deleting the repeated sentences, the final size of the pre-training dataset in AraBERT is 70 million phrases (about 24 GB), which can represent a large number of topics discussed in news from different Arab regions.

AraBERT was assessed on three tasks concerning Arabic language understanding: Named Entity Recognition, Sentiment Analysis, and Question Answering [12]. In our work, we used AraBERT in the Topic modeling task to enrich the representations and provide a significant augmentation in topic coherence by adding to neural topic models, the contextual information.

B. Neural Topic Model

ProdLDA [4] solved the problem of $p(w|\varphi, \gamma)$ (a mixture of multinomials) distribution of LDA, which consists of never making predictions that are more precise than the mixed components. This led to low-quality subjects and people's judgment is not consistent.

The way to resolve this problem is by using the weighted product of experts to convert mixed words into word level. According to the definition, the weighted product of experts can make clearer predictions than any combination of experts. ProdLDA uses the weighted product of experts to replace the mixed word hypothesis in LDA [15], which greatly improves the consistency of topics.

ProdLDA employed a VAE for LDA using a Laplace approximation for the Dirichlet distribution, which makes it possible to train a Dirichlet variational autoencoder. Moreover, this model does not directly reparametrize the Dirichlet distribution.

The only modifications to pass from LDA to ProdLDA are that ProdLDA is not normalized, and the conditional distribution of W_n is interpreted as $W_n \; |\gamma,\varphi \; \sim \text{Multinomial}(1, \sigma(\gamma\varphi))$.

For the multinomial, the relation to an expert products is very simple; if we take 2 N-dimensional multinomials parameterized by mean vectors p and q and set the natural parameters to p = σ (v) and q = σ (r), we can show that (where δ ε [0, 1]):

$$P(x|\delta\mathbf{v} + (1-\delta)\mathbf{r}) = \alpha \prod_{i=1}^{N} \sigma(\delta v_i + (1-\delta) r_i)^{x_i} \; \alpha \prod_{i=1}^{N} [v_i^{\delta} . r_i^{(1-\delta)}]^{x_i}$$

4. EXPERIMENTS AND RESULTS

We will describe in this part, our dataset (collect and preprocessing), then we will present baselines and used metrics to compare our model with baseline methods.

A. Dataset

For almost 10 years, the Application Programming Interface (API) provided by Facebook has been the primary tool for researchers to collect data on Facebook. These data contain public information about user profiles, comments and reactions to public messages. However, after the Cambridge Analytics (CA) scandal in early 2018, Facebook significantly tightened access to its API [18]. To pull data from Facebook pages, we used web scraping techniques with Python language (Version 3.8) namely Requests (Version 2.25.0) and BeautifulSoup4 (Version 4.9.3) as external libraries for automatic browsing. In our study, we were interested in Arabic posts published on Facebook by Moroccan news pages. We chose seven Facebook pages (Hespress, Medi1TV, aljarida24.ma, alakhbar.maroc, Alyaoum24, JARIDATACHCHAAB, al3omk), and collected 81 598 posts published from 04 October 2020 to 05 March 2021 (details of collected posts are shown in Table 1)

TABLE I. DATA COLLECTION

Facebook pages	Number of collected posts
Hespress	22 854
Medi1TV	18 176
aljarida24.ma	8 133
alakhbar.maroc	6 400
Alyaoum24	15 489
JARIDATACHCHAAB	1 571
al3omk	8 975
Total	**81 598**

B. Data preprocessing

We performed the common pre-processing steps in existing approaches which consist of:

- Removal of Arabic stopwords.
- Removing hyperlinks, hashes to keep only Arabic text.
- Lemmatization of words using Farasapy Lemmatizer [19] for Arabic text.
- Selecting the most frequent 2,000 words as the vocabulary

C. Word embedding

It designates a set of learning methods in NLP where vocabulary words (or phrases) are converted to numerical vectors. In our research, we used BERT word embeddings.

AraBERT is an Arabic pre-trained language model that gives a contextual embeddings, which is used to generate word embeddings. Each word is interpreted as a vector of size 768 and consequently each sentence is a list of word embeddings are extracted. The sequence of the embeddings is completed so that they have the same size.

For our AraBERTopic model, we used the pretrained bert-base-arabert Arabic embedding with 12 encoder blocks/layer, 768 hidden dimensions, 12 attention heads, and 110 M parameters. It can be found on the Google Bert model website [20].

D. LDA and ProdLDA settings

Concerning LDA and prodLDA parameters, we selected twenty topics and the top-ten words for each topic. For the number of topics (N), we tested N values from 5 to 80.

When the N value was 20, the results were good with the highest topic coherence value as shown in figure 4.

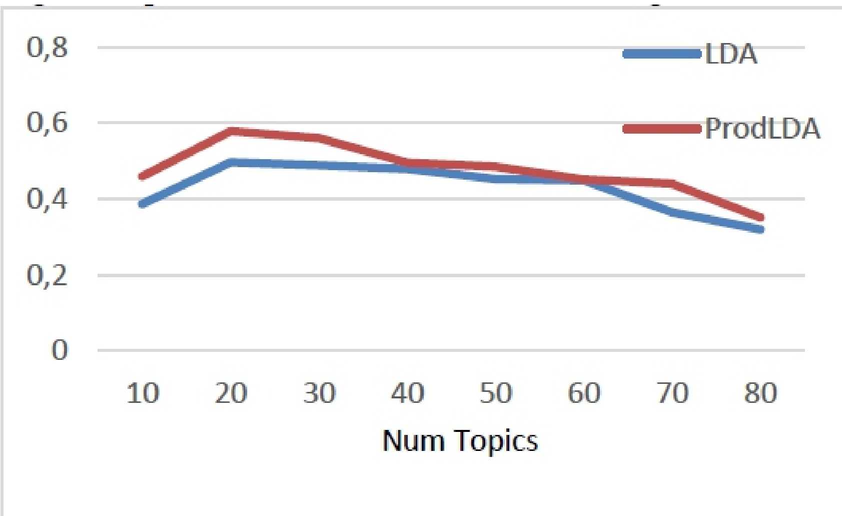


Figure 4. Choosing the optimal number of LDA and ProdLDA topics

E. Experimental Setup

In our experiments, we used the implementation of AraBERTopic in Pytorch Library (Version 1.6.0). Adam optimizer was used, with a batch size of 64 and 2e-3 for the learning rate. Our model was fine-tuned after 10 epochs over the data.

F. Baselines

We compared the two levels of our approach (feature extraction and topic modeling) to other models to prove its performance. Concerning feature extraction level, we compared Arabert Embedding with three embeddingmodels:

• **asafaya/bert-base-arabic model [5]:** This model iscontextual, it was pretrained on ~8.2 Billion words.

• **Glove [21]:** The model gives the vector representation of words using an unsupervised learning method. We choose 60% of the data set to train the model, because it captures the overall meaning of sentences in a relatively small memory.

• **Doc2vec [22]**: This model converts sentences or paragraphs to numeric vectors. In our work, we used Doc2vec Gensim implementation. To train our Doc2vec model, we used the same dataset as for the Glove model.

Concerning the second level of topic modeling, we compared ProdLDA to LDA used with different word embedding models already mentioned.

G. Metrics

To assess the performance of the topic model is usually using the following metrics:

1. Subject coherence based on NPMI algorithm,
2. Coherence value (CV) measure or topic coherence measure,
3. Confusion or Perplexity evaluation.

1) NPMI

This metric [23] gives an automatic measure of the quality of the topics to evaluate our proposed model as well as baselines. It derives from Pointwise Mutual Information (PMI) and measures the effect of one xm variable on another xn. Its formal definition is as follows:

$$\mathrm{PMI}(x_i, x_j) = \log \frac{p(x_m, x_n)}{p(x_m)p(x_m)}$$

$p(x_n)$: The likelihood that the word x_n appearing in thecorpus,

$p(x_m, x_n)$: The likelihood that the word x_m and word x_n appear together in the corpus.

We took in our experiment the top-five words of each topic, and for each word; we compute the NPMI score following this equation:

$$NPMI = \sum_{m=1}^{j} \sum_{n=m+1}^{j} \frac{PMI(x_m, x_n)}{-\log P(x_m, x_n)}$$

The topics that scored higher in NPMI are the most likely words to seem more often in the same document m than those who occasionally appeared.

2) Topic coherence

Topic coherence measure is also derived from PMI, which is used to evaluate the semantic similarity between high-resolution words of a topic. Topic coherence score is computed as follows:

$$Score_{UCI}(w_i, w_j) = \log \frac{p(w_m, w_n) + \varepsilon}{p(w_m)p(w_n)}$$

3) Perplexity

Perplexity (PPL) is a statistical measure used to assess the quality of model subject modeling. It is computed as follows:

$$Perplexity(w|z, \theta, \beta) = \exp\left(\frac{-\sum_{m=1}^{M} \sum_{n=1}^{N_m} \log p(w_{mn}|z_{mn}, \theta_m, \beta)}{\sum_{m=1}^{M} N_m}\right)$$

Where:
N_m : The number of words in the document M.
θ : Document-topic density
β : Topic-word density.

The PPL can be explained by the confusion degree of each label in the document M to the topic model. The low-PPL model can better predict the words that may seem in the document.

H. Results analysis

In our experiments, we used experimental parameter summarized in the following table 2:

TABLE II. SETTINGS OF OUR IMPLEMENTED MODEL

Parameter	Value
N Components	20
Topic Prior Mean	0.0
Topic Prior Variance	0.95
Batch size	64
Num_epochs	10
Hidden Sizes	(100, 100)
Activation	softplus
Solver or optimizer	Adam
Dropout	0.2
Learning Rate	0.002
Momentum (momentum to use for training)	0.99
Reduce On Plateau (reduce learning rate by 10x on plateau of 10 epochs)	False

1) Quantitative Evaluation

TABLE III. EVALUATION OF PERFORMANCE OF TOPIC MODELS PRODLDA AND LDA WITH DIFFERENT EMBEDDING MODELS.

Models		Metrics		
Word Embedding	Topic Model	NPMI	CV	PPL
AraBERT	*ProdLDA*	**0.553**	**0.579**	**11.25**
	LDA	0.137	0.497	20.65
Asafaya	*ProdLDA*	0.484	0.543	56.20
	LDA	0.128	0.494	61.28
Doc2Vec	*ProdLDA*	0.358	0.482	63.79
	LDA	-0.16	0.477	78.63
Glove	*ProdLDA*	0.333	0.534	86.9
	LDA	0.109	0.434	90

We calculated NPMI, topic coherence, and perplexity measure of each model to better compare their performance with different embedding models on short Arabic text. As can be seen in Figure 5, the NPMI of AraBERTopic (AraBERT + ProdLDA) is 0.553, which is higher than the other models. Overall, our model outperforms baselines methods in terms of topic coherence value, perplexity (a lower value is better), and NPMI score as shown in Table 3.

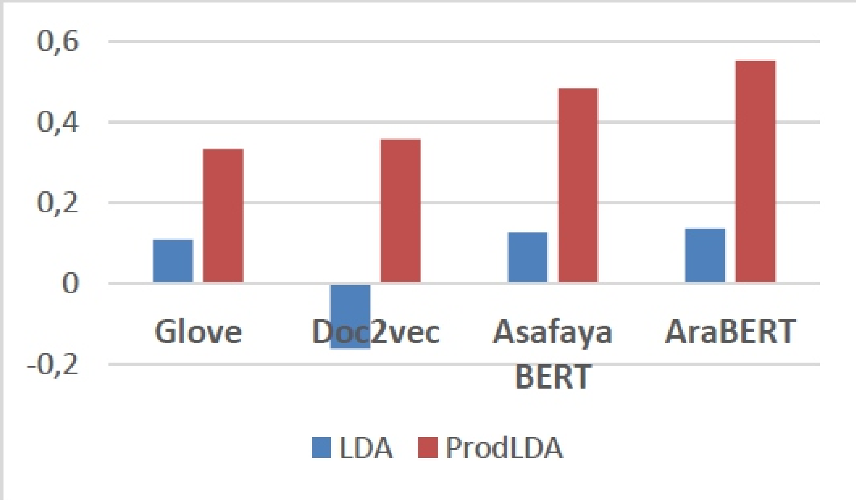


Figure 5. NPMI of different models

2) Qualitative evaluation

To prove the quality of the topics extracted by our models using ProdLDA which has the highest score within different embedding models as shown in Table 3, we display, in Table 4, examples of the top-six words of three topics from all the models. We have added the translation of each word in English.

The topics generated using AraBERTopic are more coherent than those generated by other models. In the second position, we find Asafaya's topics which sound to be coherent, but are influenced by additional mixed topics; for example, the 1st topic is about Coronavirus but it includes the term 'Geographic' which is a bit far from the topic.

Finally, we present in Figure 6 the wordcloud of the first six topics extracted by our model AraBERTopic.

TABLE IV. TOP SIX-WORDS OF THREE TOPICS EXTRACTED BY ALL THE MODELS.

Embeddings models using ProdLDA	Topics
Glove	['لقاح', 'صلاة', 'الوقاية', 'تفرض', 'كورونا', 'مناطق'] [Places, Corona, Impose, prevention, Prayer, Vaccine] ['الكركرات', 'ترامب', 'الرئيس', 'يضع', 'الملك', 'يوم'] [Day, The king, Put, President, Trump, Guergarat] ['الأمازيغية', 'السياسية', 'النواب', 'بايدن', 'أمريكا', 'حملة'] [Campaign, America, Biden, Representatives, Political, Amazigh]
Doc2vec	['الحجر', 'حملة', 'كورونا', 'الوقاية', 'المواطنين', 'سلطات'] [Authorities, Citizens, Quarantine, Campaign, Corona, prevention] ['الكركرات', 'الإمارات', 'فتح', 'علاقات', 'الملكية', 'أمريكا'] [America, property, Relation, Open, Emirates, Guergarat] ['ترامب', 'النواب', 'بايدن', 'أمريكا', 'قانون', 'مشروع'] [Bill, Law, America, Biden, Representatives, Trump]
Asafaya	['كورونا', 'الصيني', 'بريطانيا', 'لقاح', 'منظمة', 'الجغرافي'] [Geographic, Organization, Vaccin, Britain, Chinese, Corona] ['الجزائر', 'الملكية', 'القوات', 'المسلحة', 'معبر', 'موريتانيا'] [Mauritania, crossing, armed, forces, royalism, Algeria] ['الرئاسية', 'ترامب', 'الانتخابات', 'بايدن', 'الأمريكية', 'جو'] [Joe, Amirican, Biden, Elections, Trump, Presidential]

AraBERT	['حملة','لقاح','كورونا','فعالية','العالمية', ' مواجهة'] [Confrontation, global, Efficacy, Corona, Vaccine, Campaign] ['معبر','تدخل', "الجيش", "البوليساريو", 'مغاربة', "الكركرات"] [Guergarat, Moroccans, Polisario, Army, Intervention, crossing] ['الدمقراطي', "الجمهوري", 'ترامب', "الانتخابات", 'بايدن', 'الأمريكية'] [American, Biden, elections, Trump, Republican, Democrat]

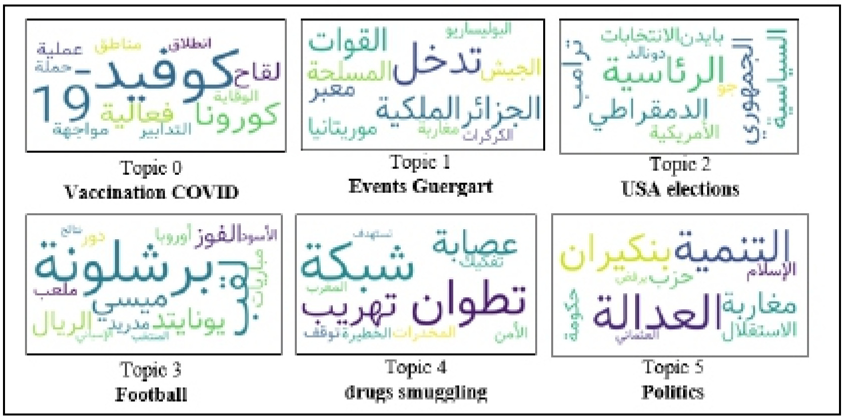


Figure 6. Topics sampled by AraBERTopic

5. CONCLUSION AND FUTURE WORK

We proposed in this paper a contextualized and neural AraBERT Topic Model (we named AraBERTopic) which integrates contextual knowledge to the neural topic model to capture more coherent and meaningful topics published in pages on the net.

For that, we collected 81 598 Arabic posts from Facebook pages of 7 Moroccan electronic press newspapers, then we preprocessed our dataset and extracted features using different embeddings models (Glove, Doc2Vec, and Asafaya – Arabic BERT). Finally, we extracted hidden topics in these pages using ProdLDA as a neural topic model and standard LDA.

To verify our contributions quantitatively, we performed experiments in terms of perplexity, topic coherence, and NPMI measures. The results proved that our proposed model can effectively capture meaningful topics and enhance the performance of topic modeling.

As part of our future work, we aim to enrich our ArabBERTopic model with new components to apply it to other tasks such as sentiment analysis using different deep learning algorithms (CNN, LSTM …).

REFERENCES

[1] ŧ Social Media users in Morocco - January 2021 ŧ. https://napoleoncat.com/stats/social-media-users-in-morocco/2021/01 (consult · le f髒r. 18, 2021).

[2] S. K. Ray, A. Ahmad, et Ch. A. Kumar, ŧReview and Implementation of Topic Modeling in Hindi ŧ, Appl. Artif. Intell., vol. 33, no 11, p. 979-1007, sept. 2019, doi: 10.1080/08839514.2019.1661576.

[3] J. Devlin, M.-W. Chang, K. Lee, et K. Toutanova, ŧBERT: Pre-training of Deep Bidirectional Transformers for Language Understanding ŧ, ArXiv181004805 Cs, mai 2019, Consult · le: janv. 14, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1810.04805.

[4] A. Srivastava et C. Sutton, ŧAutoencoding Variational Inference For Topic Models ŧ, ArXiv170301488 Stat, mars 2017, Consult · le: janv. 12, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1703.01488.

[5] A. Safaya, M. Abdullatif, et D. Yuret, ォKUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media サ, in Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona (online), déc. 2020, p. 2054-2059, Consult・ le: f雕 r. 11, 2021. [En ligne]. Disponible sur: https://www.aclweb.org/anthology/2020.semeval-1.271.
[6] H. Jelodar et al., ォLatent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey サ, Multimed. Tools Appl., vol. 78, no 11, p. 15169-15211, juin 2019, doi: 10.1007/s11042-018-6894-4.
[7] M. R. Bhat, M. A. Kundroo, T. A. Tarray, et B. Agarwal, ォDeep LDA : A new way to topic model », J. Inf. Optim. Sci., vol. 41, no 3, p. 823-834, avr. 2020, doi: 10.1080/02522667.2019.1616911.
[8] L. Liu, H. Huang, Y. Gao, X. Wei, et Y. Zhang, ォNeural Variational Correlated Topic Modeling サ, p. 11.
[9] X. Zhao, D. Wang, Z. Zhao, W. Liu, C. Lu, et F. Zhuang, ォA neural topic model with word vectors and entity vectors for short texts サ, Inf. Process. Manag., vol. 58, no 2, p. 102455, mars 2021, doi: 10.1016/j.ipm.2020.102455.
[10] J. Eisenstein, A. Ahmed, et E. P. Xing, ォSparse Additive Generative Models of Text サ, p. 8.
[11] D. Card, C. Tan, et N. A. Smith, ォNeural Models for Documents with Metadata サ, Proc. 56th Annu. Meet. Assoc. Comput. Linguist. Vol. 1 Long Pap., p. 2031-2040, 2018, doi: 10.18653/v1/P18-1189.
[12] F. Bianchi, S. Terragni, et D. Hovy, ォPre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence サ, ArXiv200403974 Cs, avr. 2020, Consult・ le: janv. 14, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/2004.03974.
[13] Q. Xie, X. Zhang, Y. Ding, et M. Song, ォMonolingual and multilingual topic analysis using LDA and BERT embeddings サ, J. Informetr., vol. 14, no 3, p. 101055, ao嗷 2020, doi: 10.1016/j.joi.2020.101055.
[14] W. Antoun, F. Baly, et H. Hajj, ォAraBERT: Transformer-based Model for Arabic Language Understanding サ, p. 7.
[15] I. A. El-khair, ォ1.5 billion words Arabic Corpus サ, p. 6.
[16] O. ElJundi, W. Antoun, N. E. Droubi, H. Hajj, W. El-Hajj, et K. Shaban, ォ hULMonA: The Universal Language Model in Arabic サ, p. 10.
[17] D. M. Blei, ォLatent Dirichlet Allocation サ, p. 30.
[18] M. Mancosu et F. Vegetti, ォ What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data サ, Soc. Media Soc., vol. 6, no 3, p. 205630512094070, juill. 2020, doi: 10.1177/2056305120940703.
[19] MagedSaeed, farasapy: A Python Wrapper for the well Farasa toolkit. .
[20] google-research/bert. Google Research, 2021.
[21] J. Pennington, R. Socher, et C. Manning, ォGloVe: Global Vectors for Word Representation サ, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, oct. 2014, p. 1532-1543, doi: 10.3115/v1/D14-1162.
[22] Q. V. Le et T. Mikolov, ォDistributed Representations of Sentences and Documents サ, ArXiv14054053 Cs, mai 2014, Consult・ le: f雕r. 11, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1405.4053.
[23] J. H. Lau, D. Newman, et T. Baldwin, ォ Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality サ, in Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, avr. 2014, p. 530-539, doi: 10.3115/v1/E14-1056.

Mrs Nassera HABBAT PhD student in Hassan II University Morocco, obtained his Master degree in Information Systems Engineering from Caddi Ayyad University, and Licence degree in Technology and Web Programming from the same University in 2015 and 2013 respectively. Her research interest are: Big data and Machine learning.

Dr. Houda ANOUN received her engineering degree in Software Engineering from ENSEIRB Bordeaux in 2003 and her PhD in computational linguistics from Bordeaux I University in 2007. And actually she is a professor in the Computer Science Department at Ecole Sup騁ieure de Technologie (Hassan II University), where she has been since 2009. Her current research interest lie in the area of IA especially deep learning, machine learning, and Big Data.

Dr. Larbi Hassouni got his engineer degree in 1983 from the “École Centrale de Marseille”. He prepared his Phd degree in 1987 at the University of Aix Marseille III in France. Among his research work, there is the development of a list unification software with LeLisp language of INRIA in order to contribute to the realization of the inference engine of an expert system for the digital circuits' diagnossis. He also developed a behavioral symbolic simulator of digital circuits using the C, FRL, and LeLisp languages in order to contribute to the development of a “formal proof” tool for correcting a design of material produced by a Hardware Design Language (HDL). Currently, his research work mainly concerns Data Sciences.

ISSN 2210-142X

A Modified Split-Radix Architecture-Based Key Scheduling Technique for Lightweight Block Ciphers

***Sohel Rana[1], Mohammod Abul Kashem[2]**
[1,2] Depart Computer Science and Engineering, Dhaka University of engineering and Technilogy, Gazipur-1700, Bangladesh
sohelresearch@gmail.com[1]

ABSTRACT

In this paper, a new key scheduling technique termed MSBK is proposed that introduces the split-radix butterfly architecture of fast Fourier transformation (FFT) in the area of cryptography. The MSBK cipher modified the butterfly architecture to fit accurately into cryptography. However, butterfly architecture has a higher avalanche effect which creates generated keys enough strong to protect information from different online and offline attacks. It also meets the standard of the Shannon properties: Confusion and diffusion. The proposed MSBK cipher consumes less power than earlier works. The memory usage and execution cycle of the MSBK technique have been evaluated on the FELICS (Fair Evaluation of lightweight Cryptographic Systems) tool that runs on the Linux operating system. The proposed MSBK technique has also been implemented in MATLAB 2021a to test the key sensitivity by the histogram, correlation graph, and entropy of several encrypted and decrypted images. The negative correlation coefficient of images encrypted by the proposed cipher indicates that the differential and statistical attacks are quite impossible for an intruder.

Keywords: Split-radix Butterfly architecture, Avalanche Effect, Lightweight cryptography, Block Ciphers, Key Scheduling, FELICS, MATLAB.

1. INTRODUCTION

The lightweight cryptography [1] is that the scaled-down version of traditional cryptography which target is to provide security for devices which resource capacity is restricted. While thinking in computer communication all people want that no unwanted person but the expected one can receive original information. To make sure this security concern we must encrypt our message in such a way that only the expected one can decipher the information. However, there is a clear trade-off between lightweight and security at the core of lightweight cryptography: that is how a decent level of security are often achieved in such sort of resource-constrained devices. In recent years, the research community has been focusing on designing cryptographic primitives which are suited to those resource-constrained devices [2]. Conventional cryptographic algorithms like RSA mostly perform well in powerful devices; therefore lightweight algorithms do not seem to be necessary for them. Embedded systems, radio frequency identification devices (RFID), Internet of Things (IoT), and sensor networks are resource constraints. So, lightweight cryptography is mainly motivated for those. Different types of cryptographic algorithms like AES [1], DES [2], PRESENT [4], etc. are used for resource-constrained devices. In most resource-limited devices, lightweight block ciphers use such design architectures which will ensure enough security while keeping execution cycles as less as possible. Most of the ciphers follow the Feistel Architecture like SIT [5], SIMON [6], Speck [6], etc. or by Substitution-Permutation Network like PRESENT [4], AES [7], etc. or by using both Architectures like DES [3], SIMON [6] to supply enough Shannon's confusion and diffusion properties in cipher text. Key scheduling in the block ciphers should perform in a secure way because the security of the ciphers depends on the secret keys which are used in every round of a block cipher.

To make round keys strong, different complex number theories like modular arithmetic, prime factorization, Euclidian algorithms, etc. are applied in key generation techniques that end in hamper the performance of resource-limited devices. A good key scheduling must have two properties; randomness to generate unique keys and high avalanche effect to ensure high key sensitivity. A single bit change in the key should change at least 50% bit in the cipher-text so that an attacker cannot easily predict a plain-text or keys through a statistical attack of cipher text. This property is regarded as avalanche effect. To implement a strong cipher, avalanche effect should be considered as one of the primary design objective.

A. Motivation

To research on cryptography is a challenging and interesting topic. Cryptography is the heart of secure data transmission. It includes complex mathematics, advanced programming, advanced number theories, etc. With the increasing usage of resource-constrained devices, lightweight block ciphers will be essential to provide security for those devices in near future. HP investigate that above seventy percent of re-source-limited devices are vulnerable to attacks [8]. It is essential to preserve a balance in between the security and performance. Since most of the cipher proposed till date is based on [3] complex number theory i.e., Modular arithmetic, Prime factorization, GCD testing algorithms etc. As conventional ciphers use complex number theory to meet (Avalanche Effect) [1] Shannon confusion and diffusion properties so these ciphers generally become computationally expensive that hinders the performance of resource limited devices. For that reason, it's becomes challenging to implement these heavy algorithms in small computing devices for ensuring security. We proposed a simple and less power consuming key generation technique based on [9] modified butterfly architecture of FFT. The average avalanche effect of MSBK technique is more than 50 % that meets the Shannon confusion and diffusion properties. On the other hand, proposed technique is computationally lightweight because MSBK uses simple logical operations like XOR and XNOR to generate output. It also ensures the non-linearity as MSBK uses random number to generate output. Thus, the proposed MSBK technique ensures enough security as well as consumes less power.

B. Contribution

In this paper, we proposed a modified butterfly architecture based key generation technique termed MSBK for lightweight block cipher to deal with the security and resource utilization challenges for lightweight devices. We just put the modified butterfly architecture into the key scheduling process to generate strong round keys. To achieve more avalanche effect that is more than 50% of output bit should be changed for a single bit change in input; this property is also called Shannon confusion; it's a non-linear transformation. It's highly recommended to design a strong cipher. My observation shows more than 53.27% avalanche effect on average for the key. It is also addressed non-linearity because MSBK uses random number to generate output. Normally this non-linear transformation is done by S-BOX [11] like in AES [3], DES [7]. We used MATLAB 2018b to test the randomness by entropy, Correlation, and histogram and key sensitivity i.e., Avalanche effect by encrypting images. The memory usage and execution cycle of the MSBK technique have been evaluated on the FELICS (Fair Evaluation of lightweight Cryptographic Systems) tool that runs on the Linux operating system. Thus, the proposed MSBK technique ensures enough security as well as consumes less power.

2. BACKGROUND LITERATURE REVIEW

Most of the modern cryptographic algorithms proposed are based on complex number theory like [1] RSA, [1] ElGamal, and SPN [3] network like AES and [1] Fiestel architecture like [3] DES. The primary focus of using these primitives is to create keys and cipher text more secure by ensuring more avalanche effect i.e., more confusion and diffusion in cipher text. The main problem with these primitives: these are computationally expensive which hampers the performance of resource limited devices if implemented on. So, to achieve more avalanche effect with lightweight mathematical operation is a difficult issue in cryptography. We proposed a neural network based technic to solve this issue.

A. Modified Butterfly architecture of FFT

The widely used low-complexity implementation of the Discrete Fourier Transform (DFT) is the FFT [10] which involved processing the signal for resource-limited devices. Butterfly architecture is the heart of FFT computing. One of the most popular FFT methods is the Cooley–Tukey algorithm [10]. This algorithm is used to compute the complex series of FFT. This algorithm works in a divide-conquer manner to split the whole DFT problem into several possible smallest DFTs. The simple FFT consists of radix-2 butterfly architecture blocks. We used the same structure of butterfly block as in FFT. We just replaced complex computation and mathematical operations by two logical operations: XOR and XNOR that play a vital role to achieve a consistent avalanche effect in generated keys.

B. Feistel Architecture

Feistel architecture is a symmetric structure to achieve higher avalanche effect in cipher text to keep safe cipher text from different attacks. It is a repetitive structure. There are several rounds and each round has same operation with different keys in Feistel Network. Input data is divided into two halves. The right half does unchanged and also is transmuted by a round function which takes a round-key as input. The left half generates output by combining with the transformed output of right half using a bitwise operation XOR. After that left half and right half are exchanged to get input for the next round. The number of round depend on cipher design.

C. Related works

The author of article [9], Eva Volna proposed a cipher that deals with using neural network in the area of cryptography. They designed an encryption algorithm based on neural network that would be practically in cryptography. They applied Backpropagation network as a supervised learning to train the fully connected feed forward neural network. They claimed that the neural network works reliably and unconditionally no errors are initiated in the outputs throughout encryption and decryption process. Weights treat as keys in the neural network based cryptosystem. But their algorithm is computationally expensive due to extra burden of training process of neural network.

In paper [14], the authors evaluate the performance and security of modern lightweight ciphers like TEA, HIGHT, KATAN and KLEIN which are instigated especially in resource-constrained devices. These ciphers are implemented on AVR Atmel ATtiny45 microcontroller to estimate performance based on their memory efficiency and energy consumption and also assessed the degree of confusion and diffusion to test avalanche effect.

In paper [2], authors proposed a cipher for lightweight devices in 2018 that consists of two core concepts genetic algorithm namely two point crossover and coin flip mutation which end result to achieve less data usage and also condensed power consumption than the earlier proposed ciphers. Authors also contributed on key scheduling process by putting non-linear bit shuffling in replace of matrix. They also test their proposed cipher on FELICS to compare execution cycle and memory usages. They analyzed their cipher based on the image encryption by MATLAB tools.

The Paper [5] presented a symmetric cipher that combined together Feistel architecture and Substitution Permutation Network (SPN) to avail the linear and non-linear transformation in cipher text. They proposed a cipher that includes: key generation and the encryption process. The key scheduling section generates 5 unique keys by taking 64 bits master key as input from the user. After initial permutation, 64-bit input is grouped into 4 blocks each of which is 16-bit data in size. Every 4 blocks are fitted as input for f-function as shown in Fig. 1. A 4x4 metric is used to transform the output of the f-function. Here the only source of nonlinearity is the usage of the matrix. The F-function consists of two P-Boxes; used in key scheduling is just the linear transformation. The F-function consists of two P-Boxes; used in key scheduling is just the linear transformation.

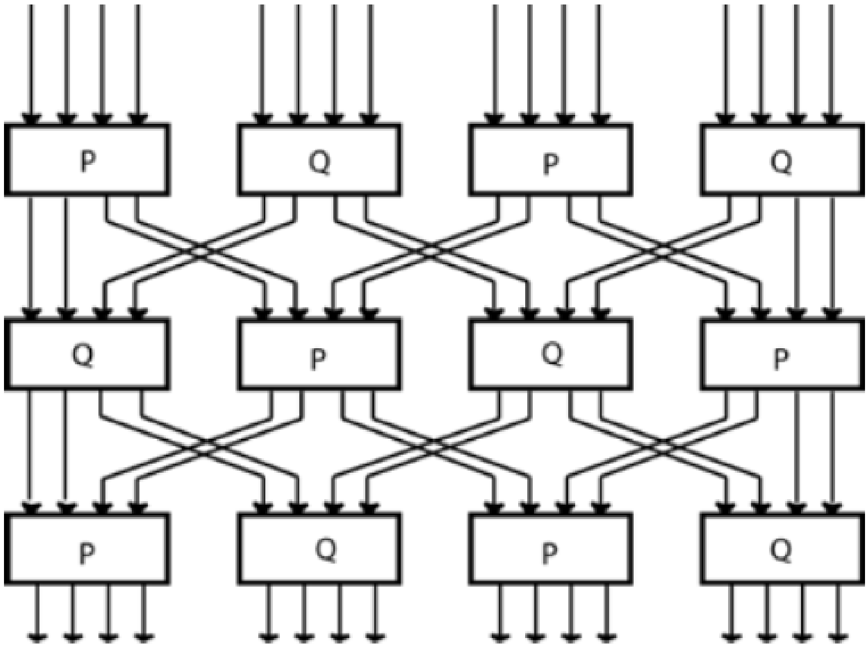


Figure 1: F-Function of SIT algorithm.

We introduced MSBK-function in replace of F-function to provide both nonlinearity and high avalanche effect which results in high key sensitivity.

3. PROPOSED APPROACH

This section describes our proposed approach that is a symmetric key block cipher of 64 bits block size. The proposed algorithm consists of Key scheduling as well as encryption process. The key expansion generates round keys which are used to encrypt plaintext with the help of encryption process. The proposed approach uses MSBK based key scheduling technique to generate five unique round keys for five round of encryption process. Thus the encryption process must be strong enough so that the intruders cannot be able to break the cipher. In the process of encryption and decryption, the most fundamental component is a key. The whole security of the encrypted cipher text depends on that key. If the key that was used to generate cipher text is compromised, the security is totally broken. The size of the round keys is 64 bits.

A. MSBK Structure

The Modified Split-radix Butterfly for Key (MSBK) structure is very simple which is based on the concept of the modified split-radix butterfly structure of FFT. This function have three layers namely input layer X={x0, x1, x2, x3}, middle layer and output layer Y = {y0, y1, y2, y3}. This function takes four 4-bit numbers as input and output has the same size as input. Two basic bitwise logical operations: XOR and XNOR are used to produce the output. The modified split-radix butterfly comprises two hidden layer operations from full butterfly architecture for first two input sequences like X0 and X2 (see Figure 2). In the split-radix architecture, first two input sequences like X0 and X2 is directly connected to output layers as shown in Figure 2. Also two random numbers R1 and R2 is applied to generate the output.

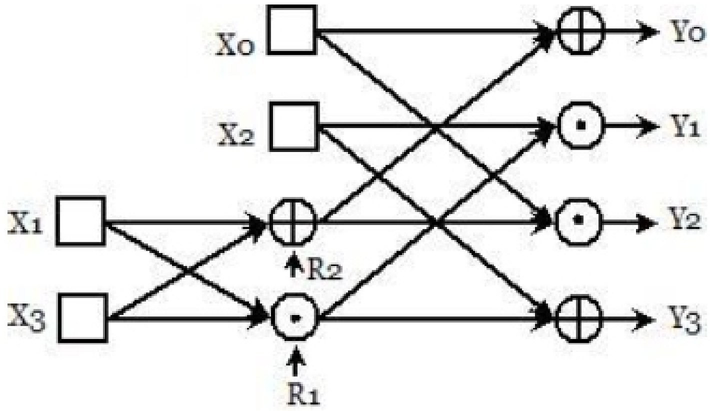


Figure 2: Internal structure of MSBK function

The pseudo random number R1 and R2 are generated by the following equation.

$$X = \sum_{i=0}^{4} x_i \qquad (3.1)$$
$$R_1 = X \bmod m \qquad (3.2)$$
$$R_2 = (X+1) \bmod m \qquad (3.3)$$

where $x_i = [x_0, x_1, x_2, x_3]$ four inputs and m is a prime number in between 2 and 16.

The pseudo random number ensures the nonlinear property of MSBK function. The output of the function is generated as the following equations.

$$Y_0 = x_0 \oplus x_1 \oplus x_3 \oplus R_2$$
$$Y_1 = x_2 \odot x_1 \odot x_3 \odot R_1$$
$$Y_2 = x_0 \odot x_1 \oplus x_3 \oplus R_2$$
$$Y_3 = x_2 \oplus x_1 \odot x_3 \odot R_1$$

Thus, output layer Y is calculated.

B. Key expansion with MSBK

The key scheduling architecture of a cipher should be as difficult as possible to prevent data from different statistical attacks like chosen cipher text, chosen plain text, differential attack etc. the sensitivity of the keys must be too high. Even if the attacker assume the key that differs only a single bit from the original key, the result of decryption with that assumed key should be as like as cipher text. To ensure higher key sensitivity, we used MSBK structure for generating key with avalanche effect more than 50%.The Figure 3 illustrates the details of proposed key expansion technique.

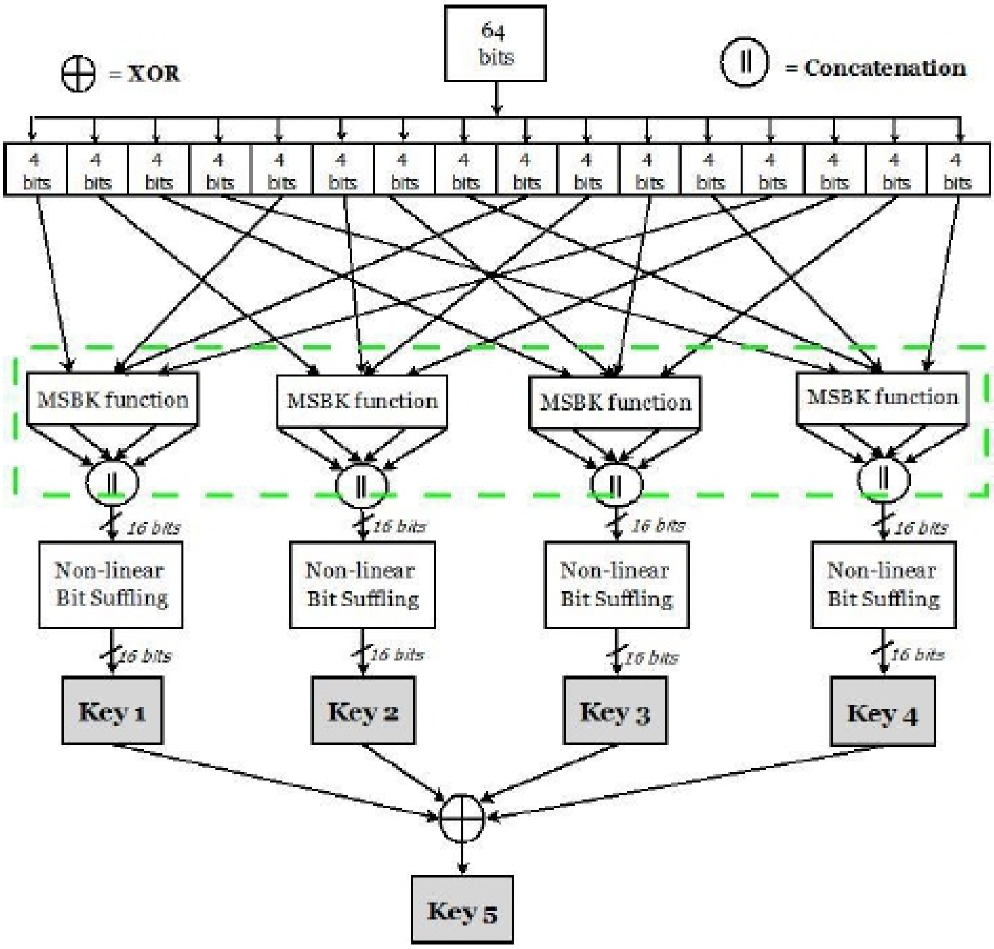


Figure 3: MSBK based Key expansion

The proposed algorithm needs 64 bit data as input to generate five unique keys for five rounds of encryption process as shown in figure 3. These 64 bits data are divided by 4 bits which generate 16 networks. Each network of 16 networks, consist of 4 bit data. The proposed technique uses 4 MSBK blocks which operate on four 4 bit (0-15) numbers. These four numbers for each MSBK blocks are obtained after performing an initial permutation of 16 segments of input key (K) as like equation (3.4).

$$MSBK_i = ||_{j=1}^{4} K_{(j-1)+i} \qquad (3.4)$$

Where i = 1 to 4 for first four round keys as shown in Figure (3). Each MSBK blocks takes input of every four segments assigned as in equation (1). To make enough diffusion (linear transformation) in generated keys, we replicated the nonlinear bit shuffling from previous research. After that the four 16-bit key (K1, K2, K3, K4) are generated which consists of the output of four bit networks. The next step is to apply XOR operation on every key K1, K2, K3, and K4 to generate the fifth key K5.

C. Avalanche effect of MSBK based key expansion structure

The proposed MSBK based key scheduling process generates round keys with higher avalanche effect. We generated quite thousands keys to evaluate the avalanche effect of MSBK based key generation approach. In best case analysis, it provides avalanche effect up to 63.49%. However, the average avalanche effect of proposed key scheduling is above 50 percent which meets the standard of Shannon confusion properties. The Table 1 shows 3 pairs of cipher text for 3 different pair of main keys while a pair of keys differs only in single bit to each other.

Table 1: Avalanche effect of MSBK based approach

SL	Main Keys(64 bits)	Cipher text (64 bits)	No. of bit changed	Avalanc-he Effect
1	0x1000000000000001	0xd9bca81c1d41fe01	36	56.25
	0x1000000000000004	0x13c10b4a82272460		
2	0x5555555555555554	0x52f4f57878257569	40	63.49
	0x5555555555555555	0x9c2f0f7996108fbe		
3	0xabababababababab	0x7a099c9711c641b4	33	51.56
	0xabababababababaa	0x211f22faab8722ba		

*Here, Plain text used is 0x ab cd 12 34 87 65 01 35

D. Non-Linear Bit Shuffling

We used the nonlinear bit-shuffling as reported in the literature [2]. In the nonlinear bit-shuffling, a concatenated 16-bit is transferred to each block of non-linear bit shuffling. After that, a random number is computed from the input of 16-bit data which is again logically combined with 16-bit input by performing a bitwise XOR operation. The output of the XOR operation is transferred to bit shuffling as well as to the perfect shuffling block sequentially to create enough diffusion in generated keys. The above figure shows the flow of non-linear bit shuffling. A concatenated 16-bit input enters into each non-linear bit shuffling block form MSBK block. Taking that 16-bit data as a seed a pseudo random number is generated using linear feedback shift register. Then the two is XORed. The result is transferred to the bit shuffling block. The bit shuffling blocks perform an in-place conventional permutation.

E. Encryption process

We used the encryption algorithm which was proposed by [2]. The encryption process consists of Feistel architecture with G-function [2] which is based on the concept of two operators of genetic algorithms: mutation and crossover. Fig. 4 illustrates the flow of operation for a single among five rounds.

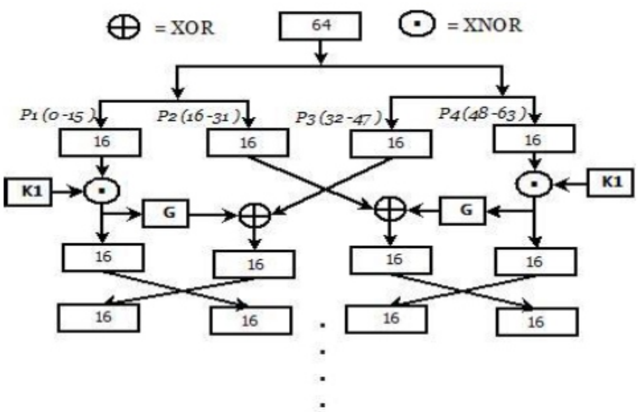


Figure 4: One of the rounds encryption process.

$$Ro_{i,j} = \begin{cases} Px_{i,j} \odot K_i & ; \quad j = 1\ and\ 4 \\ Px_{i,j+1} \oplus EG_{li} & ; \quad j = 2 \\ Px_{i,j-1} \oplus EG_{ri} & ; \quad j = 3 \end{cases} \quad (3.5)$$

The results of the final round are concatenated to obtain Cipher Text (Ct).

The 64-bit message is equally split into four 16 bit segments. According to the Feistel structure, swapping, XOR, XNOR operations are performed among the split blocks to increase the avalanche effect in cipher text. An XNOR operation is performed between round key and left as well as rightmost blocks separately. Then the output of the XNOR operation is feed to G-function as input. The 4th block and output of the left G-function are again XORed and the 2nd block and output of the right G-function are XORed as well. After that, a swapping operation is executed among the 4 blocks except for the last round. The equation (3.5) represents the process of how a single rounds encrypt plaintext into cipher text. Finally, every four blocks are combined together to generate a block of 64-bit cipher text. The decryption process is the opposite of the encryption process. This time last key is used first.

F. G-Function

We replicated the G-function from the earlier research work. The G-function [2] is based on the concept of two operators of genetic algorithms: mutation and crossover. This function takes 16 bits as input and first, split equally into 2 eight-bit segments. Both two 8 bit data are transformed by using a substitution box that is called S-Box. After performing a two-point crossover to both outputs of S-Box, a coin flip mutation operation occurs. Finally, the 16-bit output is generated.

4. RESULT AND DISCUSSION

The proposed approach is implemented in C language. We used CodeBlocks as IDE. The coding was free of any machine specification. We also used a benchmark tool termed FELICS to measure memory usage and execution cycles that runs on Linux Ubuntu. The FELICS tool is open access and free to install. We also implemented our proposed MSBK cipher in MATLAB 2021a to evaluate the security strength of keys by encrypting images.

A. Evaluation Parameters

The security strength of the proposed algorithm is tested based on key sensitivity, execution cycles, histogram and correlation. We also assessed the memory utilization and execution cycles for key generation, encryption and decryption of this algorithm.

B. FELICS implementation

The MSBK cipher is also implemented in a benchmark tool termed FELICS to measure memory usage and execution cycles .The FELICS [15] provides a command-line interface like GCC (GNU Compiler Collection) to test and build any lightweight cryptographic code. They provide documentation to facilitate the implementation as shown in Figure 5. We can compile our implementation and test whether ours is runnable in FELICS or not. It provides three scenarios against which we can test our code.

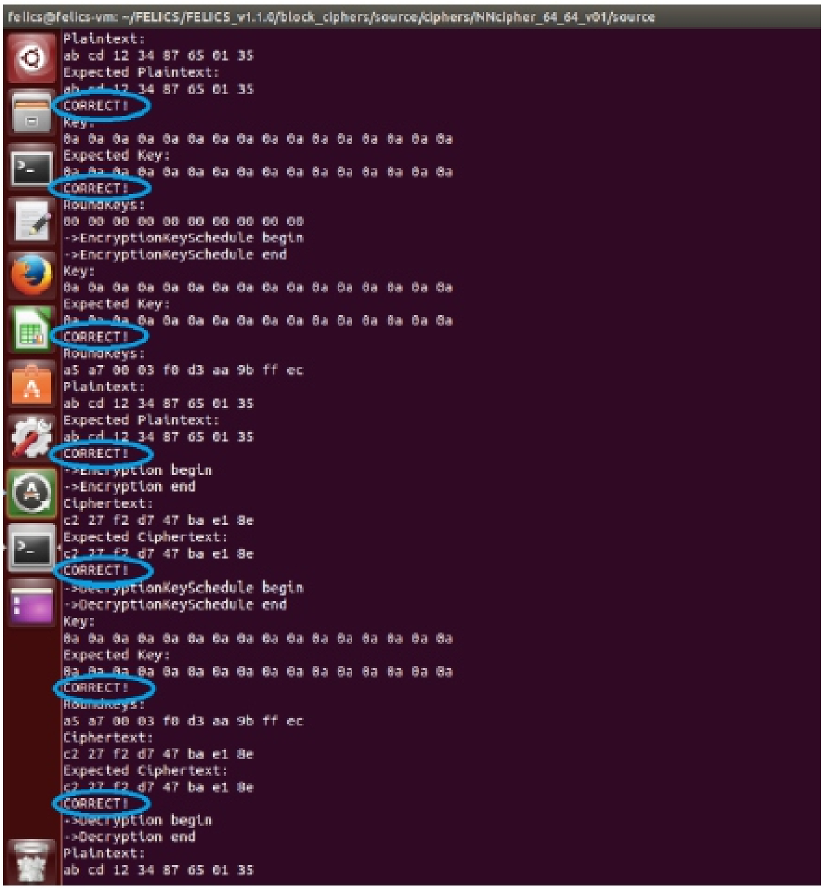


Figure 5: Testing the implementation of the proposed approach on FELICS.

The simulation of the algorithm is performed by an open source benchmark tool for lightweight cryptography named FELICS. It is used for the purpose of performance evaluation on different platforms (such as AVR, MSP, ARM and PC). It can measure execution cycles, RAM footprint and binary code size. It can easily compare the new cipher with previous ciphers. Table IV shows a comparative results of different ciphers for AVR architecture. It can be seen that among the methods considered, the proposed method has the lowest cycles for key generation, encryption and decryption.

Table 2: Comparison Of Different Lightweight Algorithms On Avr Architecture

CIPHER	Device	Block size	Key size	CODE SIZE	RAM	Cycles (Key generation)	Cycles (encryption)	Cycles (decryption)	Total execution Cycles
AES[7]	AVR	128	128	23090	720	3274	5423	5388	
HIGHT[16]	AVR	64	128	13476	288	1412	3376	3401	14085
LEA[14]	AVR	128	128	3700	432	4290	3723	3784	8189
PRESENT[4]	AVR	64	80	1738	274	2570	7447	7422	11797
Simon[6]	AVR	64	96	1370	188	2991	1980	1925	17439
Speck[6]	AVR	64	96	2552	124	1509	1179	1411	6896
SIT[5]	AVR	64	64	826	22	2130	876	851	4099
G-cipher[2]	AVR	64	64	1228	34	1630	792	789	3857
MSBK cipher	AVR	64	64	1228	34	1416	792	789	2997

Fig. 6 presents bar chart comparisons among various reported ciphers with proposed approach. For each ciphers, bar chart shows the status of required clock cycles to generate keys, cipher text from plaintext, plaintext from cipher text, as well as overall execution cycles. The chart shown in figure 6 clearly demonstrates that the proposed cipher executes in less clock cycles, improving over the other reported ciphers.

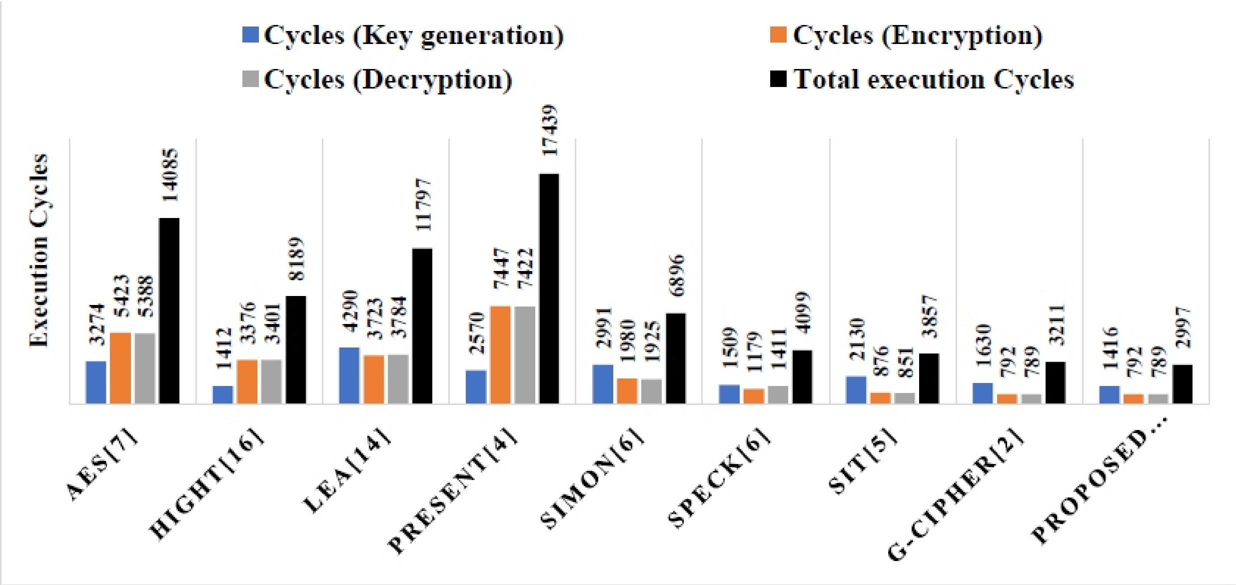


Figure 6: Execution Cycle Comparison for Hardware Implementation

C. MATLAB implementation

The MSBK cipher is also demonstrated in MATLAB® which encrypts an image and then decrypts the image with the correct key for a visual observation key sensitivity. After that the images are decrypted by using a wrong key with only one-bit difference from the original key. This is also a test for the avalanche effect of the keys. The ciphertext is not recognizable though only one bit changes in the original keys. Fig. 7 shows that for MSBK cipher, the encrypted images can only be decrypted with the correct key

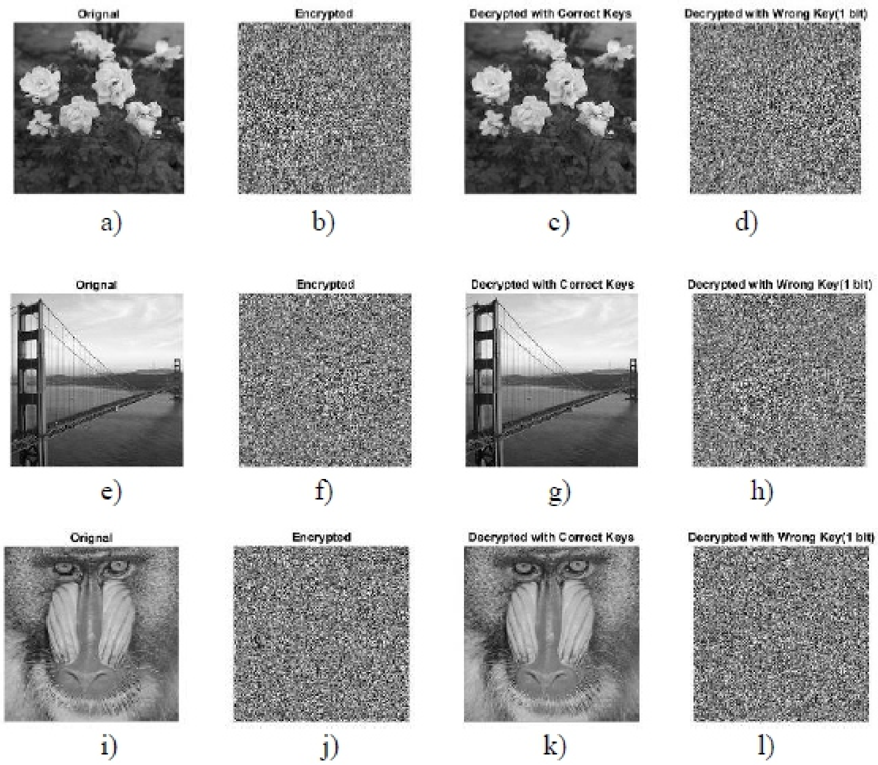


Figure 7: Statistical Analysis of Key Sensitivity

Fig. 8 presents the histograms of original images: a) Flower, e) Bridge, i) Baboon and encrypted images: b) Flower, f) Bridge, j) Baboon. The vertical line refers to the number of pixels, and the horizontal line refers to the intensity of image. Cipher images show the uniform distribution in histogram that indicates the security strength of the proposed cipher. Hence, the statistical attacks like chosen plaintext attack, chosen cipher text attacks are not vulnerable to this cipher.

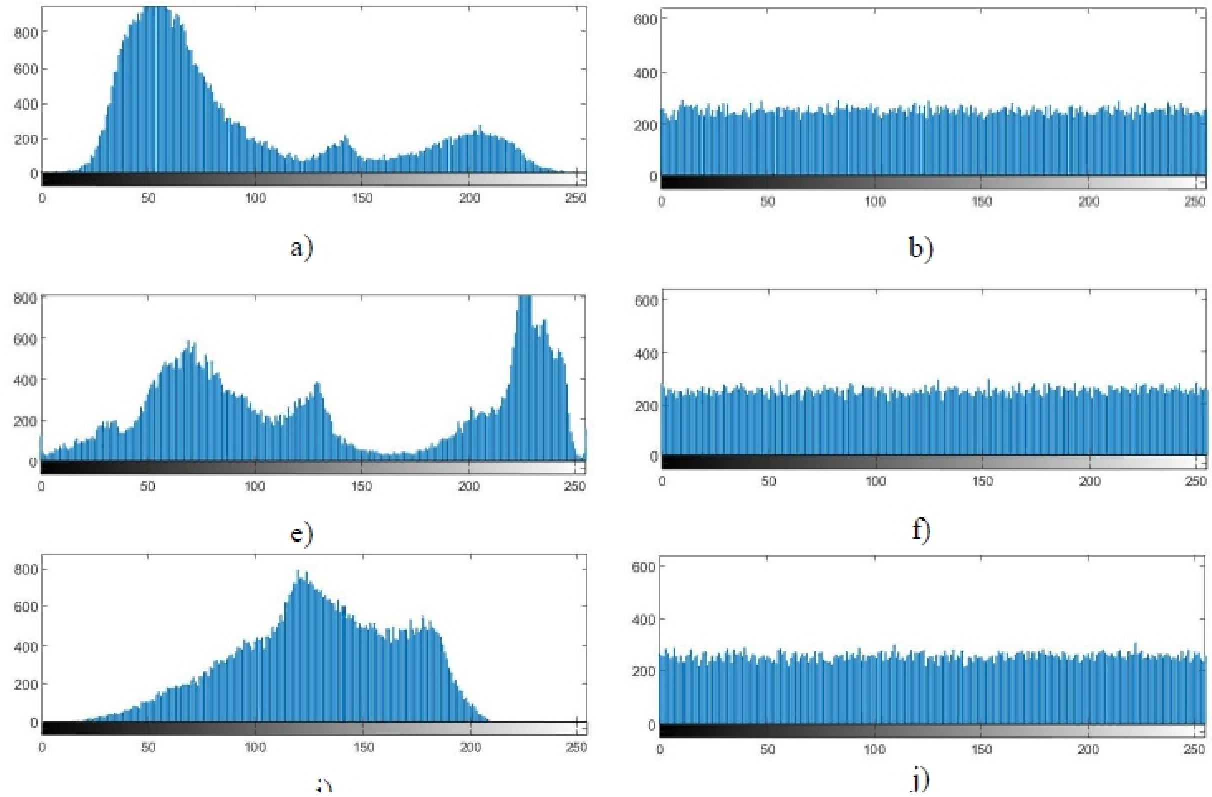
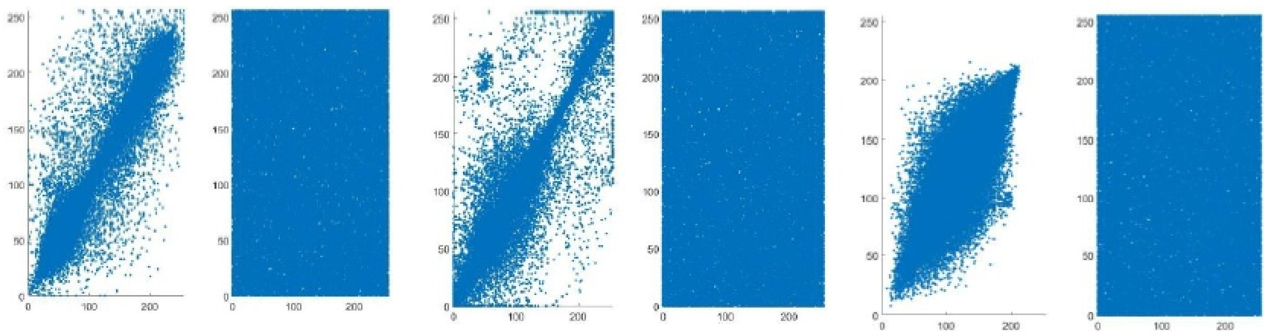


Figure 8: The histogram of original images and decrypted images

Figure 9 shows the correlation graph of the original image and the encrypted images .The correlation graph of plain image shows linear relationship that is higher positive correlated value. However, the correlation graph of cipher image shows highly randomness that is the negative values. Hence, the negative correlation values of encrypted images indicates the strong security strength of proposed cipher.



Correlations for (a) encrypted and (b) decrypted image

Correlations for (e) encrypted and (f) decrypted image

Correlations for (i) encrypted and (j) decrypted image

Figure 9: The correlation graph of original and decrypted images

The below Table 3 indicates the security strength analysis of the proposed technique using (a)Flower, (e)Bridge, (i)Baboon images correlation coefficients. The correlation coefficient of plain images shows higher positive correlated value. However, the correlation coefficient of cipher images shows highly randomness that is the negative values. Hence, the negative correlation values of encrypted images indicate the strong security strength of proposed cipher. Also number of pixel change rate (NPCR) is almost 100% that indicates the strong security of the cipher.

Table 3. Correlation Coefficients Of Original Image And Encrypted Image

Images		Correlation coefficients	UACI	NPCR (%)
Flower (a)	Original	0.9569	23.1466	99.5987
	encrypted	-0.0005		
Bridge (e)	Original	0.9626	14.7873	99.5941
	Encrypted	-0.0044		
Baboon (j)	Original	0.8198	13.3488	99.5575
	encrypted	-0.0015		

The score of UACI for all tested images meet the security standard as shown in Table 3. So, the security strength of the proposed cipher is as strong to prevent online and offline attacks.

Table 4: Results of Information Entropy

		Bridge	Baboon	Flower
Entropy H(S)	Encrypted image	7.9976	7.9972	7.9974
	Original image	7.5856	7.2316	7.2658

The typical result of entropy is 8 that relates for the real randomness. Here, proposed approach gives almost 8 of entropy for three different images as shown in Table 4.

D. Energy Consumption

For calculating the total power consumed by an algorithm on a particular devices, first we need the execution cycle of that algorithm. By using the following equation (4.1), we can compute the power consumption of an algorithm on a particular device:

$$\text{Power } E = I * V_{cc} * \text{ N} * \text{Ʈ} \quad (4.1)$$

Where, V_{cc} denotes the operating voltage and I indicates operating current used up for T seconds (unit in Ampere). Ʈ refers to the clock period as well as N indicates the required number of execution cycle. If f be the operating frequency of the particular device in Hertz then we can calculate the time period of the particular device that is Ʈ $= {}^{1}/_{f}\ sec/cycle$.

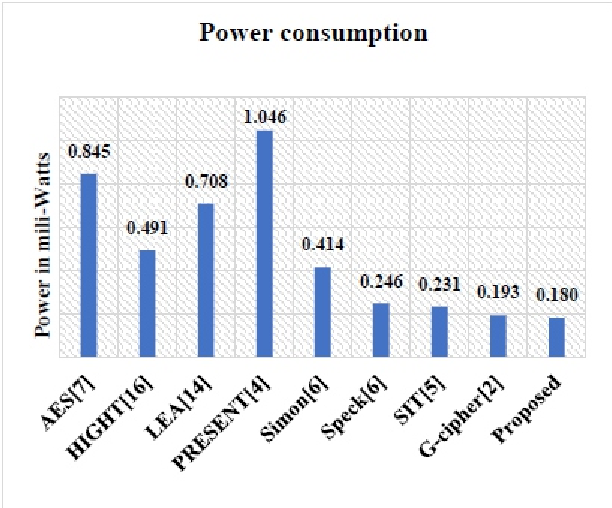


Figure 10: Energy consumption comparison of ciphers

According to the [19] absolute maximum rating (AMR) of dataset, usually maximum operating voltage of Atmel Atmega88/168 is 6V, Maximum current is 200mA and operates at 20 MHz. Figure 10 demonstrates the of energy consumption of different reported ciphers along with the proposed MBFK cipher. The bat chart shows that proposed approach consume less power than that of others. It can be seen that for the algorithms considered, the proposed algorithm has the lowest power consumption.

5. CONCLUSION

The security, as well as performance of resource-limited devices, is an important issue. For this purpose, a lightweight cryptographic algorithm using butterfly architecture of FFT is proposed in this paper. This MSBK algorithm has lower key generation cycles and less power consumption than the existing ciphers. The histogram and the correlation plots indicate that the MSBK cipher can reliably encrypt images. Moreover, the key sensitivity results indicate that for MSBK cipher, encrypted images can only be successfully decrypted using the actual key. Hence, the proposed cipher will be an excellent solution of security for those devices that are resource-limited. We intend to perform more mathematical analysis on our proposed algorithm to investigate its security strength as future work.

Ethical disclosures

Conflict of interest: Authors declare that there is no conflict of interest of this study.

REFERENCES

[1] William Stallings "Cryptography And Network Security Principles And Practices", Fourth Edition, Publisher: Prentice Hall, November 16, 2005

[2] Sohel Rana, Saddam Hossain, Hasan Imam Shoun And Dr. Mohammod Abul Kashem, "An Effective Lightweight Cryptographic Algorithm To Secure Resource-Constrained Devices" International Journal Of Advanced Computer Science And Applications (IJACSA), 9(11), 2018. DOI: 10.14569/IJACSA.2018.091137

[3] Bart Preneel, "Understanding Cryptography, A Textbook For Students And Practitioners" ISBN 978-3-642-04100-6 E-ISBN 978-3-642-04101-3, DOI 10.1007/978-3-642-04101-3, Springer Heidelberg Dordrecht London New York, ACM Computing Classification (1998): E.3, K.4.4, K.6.5

[4] Kostas Papapagiannopoulos."High Throughput In Slices: The Case Of PRESENT, PRINCE And KATAN64 Ciphers". Radboud University Nijmegen, Department Of Digital Security.2016

[5] Muhammad Usman, Irfan Ahmed, M. Imran Aslam, Shujaat Khanand Usman Ali Shah. "SIT: A Lightweight Encryption Algorithm For Secure Internet Of Things" Iqra University, Defence View And Department Of Electronic Engineering (IJACSA) International Journal Of Advanced Computer Science And Applications, Vol. 8, No. 1, 2017.

[6] Ray Beaulieu, Douglas Shors, Jason Smith, Stefan Treatman-Clark, Bryan Weeks, And Louis Wingers."The Simon And Speck Block Ciphers On AVR 8-Bit Microcontrollers". National Security Agency.2014

[7] Federal Information Processing Standards Publication 197," Announcing The Advanced Encryption Standard (AES),NIST, November 26, 2001

[8] S. A. Kumar, T. Vealey, And H. Srivastava, "Security In Internet Of Things: Challenges, Solutions And Future Directions", In 2016 49th Hawaii International Conference On System Sciences (HICSS), IEEE, 2016, Pp.5772-5781

[9] Eva Volna , Martin Kotyrba,Vaclav Kocian, Michal Janosek, CRYPTOGRAPHY BASED ON NEURAL NETWORK, Proceedings 26th European Conference On Modelling And Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring

[10] Guilherme Ferreira Guilherme Paim Leandro M. G. Rocha Gustavo M. Santana Renato H. Neuenfeld Eduardo A. C. Costa Sergio Bampi, " Low‐power fast Fourier transform hardware architecture combining a split‐radix butterfly and efficient adder compressors" https://doi.org/10.1049/cdt2.12015

[11] Sohel Rana, Wadud, Ali Azgar, Dr. M Abul Kashem, "A Survey Paper Of Lightweight Block Ciphers Based On Their Different Design Architectures And Performance Metrics" International Journal Of Computer Engineering And Information Technology, June 2019, Volume 11, Issue 6.

[12] P. Wang, Professor S. Chaudhry, S. Li, T. Tryfonas And H. Li, "The Internet Of Things: A Security Point Of View", Internet Research, Vol. 26, No. 2, Pp. 337-359, 2016

[13] S. Wang, Z. Zhang, Z. Ye, X. Wang, X. Lin, And S. Chen, "Application Of Environmental Internet Of Things On Water Quality Management Of Urban Scenic River", International Journal Of Sustainable Development & World Ecology, Vol. 20, No3, Pp. 216-222, 2013

[14] Vikash Kumar Jha," Cryptanalysis Of Lightweight Block Ciphers" Aalto University School Of Science Degree Programme Of Computer Science And Engineering, Master's Thesis, November 18, 2011

[15] D. Dinu, A. Biryukov, J. Großschädl, D. Khovratovich, Y. L. Corre, L. Perrin, "FELICS – Fair Evaluation Of Lightweight Cryptographic Systems", University Of Luxembourg, July 2015

[16] Bohun Kim,Junghoon Cho,Byungjun Choi,Jongsun Park,And Hwajeong Seohindawi."Compact Implementations Of HIGHT Block Cipher On Iot Platforms. Hindawi, Security And Communication Networks, Volume 2019, Article ID 5323578.

[17] Tope Komal, Rane Ashutosh, Rahate Roshan, S.M.Nalawade, Encryption And Decryption Using Artificial Neural Network, International Advanced Research Journal In Science, Engineering And Technology, Vol. 2, Issue 4, April 2015,

[18] Kerry A. Mckay, Larry Bassham, Meltem Sönmez Turan, Nicky Mouha. Report On Lightweight Cryptography, National Institute Of Standards And Technology,USA, March 2017.

[19] Utsav Banerjee, Lisa Ho, And Skanda Koppula."Power-Based Side-Channel Attack For AES Key Extraction On The Atmega328 Microcontroller". Massachusetts Institute Of Technology. Researchgate. December 2015

Sohel Rana (1st Author): Studying M.Sc. in ICT from BUET, Dhaka, Bangladesh. Received B.Sc. in CSE from DUET, Gazipur, Bangladesh. Now working in BUBT as a Lecturer from 2018. Interested research areas: Cryptography, Network Security etc.

Dr.Mohammod Abul Kashem (2nd Author): He started his career as an Assistant Professor at the Department of CSE, Dhaka University of Engineering & Technology (DUET) in the year 2003. He joined the Department after completion of his B.Sc. and M.Sc.Engg. Degrees from State University "Lvivska Polytechnica", Ukraine in 1996 and 1997 respectively. In 2001, he earned Ph.D. in Control Systems and Processes from National University "Lviv Politechnic" Ukraine. Subsequently, Dr. Kashem completed his Post Doctorate fellowship from University Lumiera Lyon2, France. (Erusmus Mundas Scholarship, European Commission), 2016 and He was appointed as professor at the CSE Department of DUET in the year of 2013.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Note