المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

# Predicting Kurdish EFL University Learners' Oral Reading Fluency Using Support Vector Machine

**Araz Bashar Mohammed Ali**
**Asst. Lecturer in Basic Education Department, Duhok University, Kurdistan Region- Iraq**
**Email: araz.mohammedali@uod.ac**

**Jwan Abdulkhaliq Mohammed Ali**
**Lecturer in Computer Science Department, Duhok University, Kurdistan Region- Iraq**
**Email: jowan.abdulkhaliq@uod.ac**

**Amera Ismail Melhum**
**Asst. Prof in Computer Science Department, Duhok University, Kurdistan Region- Iraq**
**amera_melhum@uod.ac**

## ABSTRACT

Investigating learners' English Oral Reading Fluency (ORF) in the contexts where English is used as a foreign language (EFL) or second language (ESL) has recently become a trending subject. This study was carried out to predict the ORF of 100 Kurdish EFL university students of the English language department, college of Basic Education, university of Duhok, Iraq in 2020 during covid-19 using the support vector machine (SVM) technique. This technique is one of the supervised machine learning techniques, and it is considered the most powerful algorithm in machine learning in terms of high accuracy; therefore, it was employed in this study. Participants' ORF was measured by two experienced human raters using the four dimensions of the Multidimensional Fluency Scale (MDFS) including expression & volume, phrasing, smoothness, and pacing, which were used as the input variables to predict the ORF as an output. Six kernels of the SVM were used in the prediction process. The results indicated that the highest accuracy of testing result was obtained on the use of SVM Linear kernel with a value of 96.2%. Confusion matrix was utilized to assess the outcomes of data classification. The results of precision, recall, and F1-score were the highest for the SVM linear kernel and their values were the same for all performance metrics with a value of 96.1%. Accordingly, it can be concluded that the performance of the SVM is considerably accurate in predicting the oral reading fluency.

**Keywords:** Oral Reading Fluency, Support Vector Machine, MDFS, Machine Learning.

 **DOI: https://doi.org/10.33193/IJoHSS.39.2022.492**

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

## I. INTRODUCTION

Researchers have defined Oral Reading Fluency (ORF) in many different ways. According to Schreiber (1980), reading fluency is being capable of reading a nontechnical text fluently without using the so much effort and being capable of maintaining a reasonable automaticity. This process does not require decoding only words; it requires the reader to be able to read meaningful sequential group of words together. Distinctively, Rasinski (2004) defines ORF as “the ability to read a text both orally and silently with correct speed, accuracy, and expression” (p. 1). He also explains that struggled readers are those who are able to maintain the accuracy of reading words; however, face challenges reading difficult words as they make frequent pauses. Above all, such readers are not able to consider punctuation marks and cannot read with proper expression.

Likewise, Grabe (2010) defines ORF as “the ability to read rapidly with ease and accuracy, and to read with appropriate expression and phrasing. It involves a long incremental process and text comprehension is the expected outcome” (p. 72). A more distinct definition is given by Nation (2014), he states that ORF is “the ability to process language receptively and productively at a reasonable speed” (p.11).

Samuels (2012) illustrates specific indicators for certain dimensions of ORF. He argues that having a good reading pace or speed and maintaining good word recognition accuracy are indicators of developing reading accuracy, while maintaining good expressivity is an indicator of developing reading fluency and comprehension.

Bailly et al. (2022) highlight the fact that assessing only reading rate or speed could not produce a well understanding of phrasing, expression and comprehension; it only produces a well understanding of reading accuracy and automaticity. They believe that using such assessment with children makes them assume that a good reader is someone who reads fast; as a result, listeners and children’s reading comprehension is negatively affected. Therefore, most of the educational institutions and researchers shifted their focus from investigating reading accuracy onto investigating the features of prosodic reading as a strong indicator of reading fluency and comprehension.

Measuring reading fluency or prosodic reading with reliable scales has been a persistent requirement for researchers. The work of Morrison and Wilcox (2020) presents three scales that have been popular among researchers. The first one is the National Assessment of Educational Progress (NAEP), which is a holistic scale that measures reading rate, accuracy and fluency (prosody); it has been used since the 1980s. The second is the Multidimensional Fluency Scale (MDFS), which is an objective and highly reliable scale (0.92 – 0.98) that measures expression and volume, phrasing, smoothness and pacing, and has been used since the late 1990s. This scale was modified in 2003 to include four dimensions instead of three. Finally, the Comprehensive Oral Reading Fluency Scale (CORFS), which was developed in 2013, lacks reliability; it measures accuracy, intonation and pausing.

Assessing readers’ oral reading fluency is significant because it can be used in the following areas: (1) to compare it to other anticipated standards; (2) to enhance learners’ motivation, especially when they receive encouraging feedback; (3) to facilitate the reading fluency longitudinal analysis; (4) to highlight effective reading


DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

strategies by comparing different kinds of reading strategies; (5) to investigate its relationship with reading comprehension; and (6) to be used as an indicator of readers' level of reading comprehension when reading a text (Ardington et al., 2021).

Until now, several researchers have focused their efforts on assessing oral reading fluency using computer software program. However, a few studies on predicting and classifying oral reading fluency using machine learning algorithms are discussed. Machine learning indicates the algorithmic procedures which can act on datasets to generate algorithms that categorize, cluster, or find patterns in new datasets. A machine learning approach is a linear regression from one independent parameter to estimate a dependent parameter. Many complicated statistical models and approaches, such as logistic regression and decision trees, are used in more recent machine learning algorithms, as are others such as Bayesian models, containing Hidden Markov Models (HMM), which are extensively employed in sequential pattern recognition. Recently, Deep Neural Networks (DNN) and Support Vector Machines (SVMs) are the most two statistical models used for classification (Sen & Mehtab, 2021). The support vector machine (SVM) method has shown excellent performance to solve classification issues in a variety of domains. Many recent researches have found that SVM (support vector machines) can generally provide better performance than other data classification techniques in respect of classification accuracy (Pisner & Schnyer, 2020). Bolaños et al. (2013) have used speech recognition and machine learning techniques (SVM) to achieve automated oral reading fluency for children. Another research applied different algorithms of machine learning to predict reading fluency of 1$^{st}$ graders with reading disabilities; however, SVMs appeared to be the most powerful classifier in terms of high accuracy so the study focused on the SVM as classifier for its research (Varol, 2009). The purpose of Ja'afar et al. (2021) study was to determine the efficacy of employing the Computer Assisted Repeated Reading (CARR) approach to improve Form One rural pupils' oral reading fluency. A study by Bailly et al. (2022) presented a computer approach for calculating multidimensional subjective assessments of young readers' reading proficiency using speech-based objective data. The method of Beck et al. (2004) for measuring student reading competency involves using data collected by a computer tutor during contacts with a student to estimate his performance on a human-administered test of ORF. Cheng (2018) discussed the real-time scoring methodology for a self-administered oral reading evaluation on mobile devices to test the components of children's ORF skills. The aim of the proposed study is that the authors want to answer whether ORF among university students can be predicted with high accuracy using machine learning method (SVM).

## II. MATERIAL AND METHODS

### A. Data Collection

Data were collected from 100 Kurdish EFL 1$^{st}$ year university students of the English language department, College of Basic Education, University of Duhok in 2020.

 DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

Participants were aged between 18 – 22 including (number of males =35) and (number of females= 65). Participants were asked to video record themselves while reading out loud a text entitled “to marry or not to marry” (See appendix ) selected from the book of Cover-to-Cover1 by Day and Yamanaka (2007). The book is designed to develop students’ oral reading fluency and comprehension using extensive and intensive reading passages that are amusing, relevant to students’ interests and of moderate length. The selected text was of moderate length consisting of 419 words. It consisted of six typical paragraphs which contained a variety of sentence structures including simple, compound and complex sentences. The passage was selected to match participants’ interest and English language level.

B. Human Scoring

After selecting the reading passage for the oral reading test, participants were first briefly introduced to the four dimensions of Rasinski’s (2004) MDFS model (volume and expression, phrasing, smoothness and pacing), see Fig. 1. Due to the Coronavirus pandemic in 2020 and the successive lockdowns of the educational institutions and other facilities, participants were asked to video record themselves reading out load from the selected text and submit their video recordings via their own digital e-learning portal platform Moodle. Then, video recordings were collected and analyzed by two experienced university teachers (raters) using the MDFS model which classifies readers into four different categories. Raters listened to each recording individually and rated the four dimensions of the scale separately. Later on, raters compared their ratings, and it was found that both ratings were almost similar. As seen on the scale, dimensions are divided into four categories; each is given between 1 – 4 grades (1 representing the lowest and 4 representing the highest). Accordingly, and for each dimension, participants with low reading skills were given 1 grade and those who had moderate reading skills were given 2 grades. Meanwhile, those readers who had good reading skills were given 3 grades, and eventually those who had excellent reading skills were 4 grades.


DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **Expression and Volume** | Reads in a quiet voice as if to get words out. The reading does not sound natural like talking to a friend. | Reads in a quiet voice. The reading sounds natural in part of the text, but the reader does not always sound like they are talking to a friend. | Reads with volume and expression. However, sometimes the reader slips into expressionless reading and does not sound like they are talking to a friend. | Reads with varied volume and expression. The reader sounds like they are talking to a friend with their voice matching the interpretation of the passage. |
| **Phrasing** | Reads word-by-word in a monotone voice. | Reads in two or three word phrases, not adhering to punctuation, stress and intonation. | Reads with a mixture of run-ons, mid sentence pauses for breath, and some choppiness. There is reasonable stress and intonation. | Reads with good phrasing; adhering to punctuation, stress and intonation. |
| **Smoothness** | Frequently hesitates while reading, sounds out words, and repeats words or phrases. The reader makes multiple attempts to read the same passage. | Reads with extended pauses or hesitations. The reader has many "rough spots." | Reads with occasional breaks in rhythm. The reader has difficulty with specific words and/or sentence structures. | Reads smoothly with some breaks, but self-corrects with difficult words and/ or sentence structures. |
| **Pace** | Reads slowly and laboriously. | Reads moderately slowly. | Reads fast and slow throughout reading. | Reads at a conversational pace throughout the reading. |

Fig. 1. Multidimensional Fluency Scale (MDFS)

C. Support Vector Machine

SVMs are group of connected supervised learning techniques used for regression and classification. They are a member of the generalized linear classification family. SVM has a special characteristic which can simultaneously decrease the empirical classification error while increasing the geometric margin. As a result, SVM is also known as Maximum Margin Classifiers. SVM depends on the Minimization of structural risk. SVM maps the input vector to a higher dimensional-space in which a maximum separation hyperplane is generated (Panup et al., 2022). On either side of the hyperplane that segregates the data, two parallel hyperplanes are built. The function of separation hyperplane is to increase the space between the two hyperplanes. It is assumed that the classifier's generalization error is better when the distance between the two parallel hyperplanes increases. we use the following form of data points:

$\{(x_1, y_{1)},(x_2, y_2), (x_3, y_3), (x_4, y_4)\ldots\ldots\ldots\ldots(x_n, y_n)\}$.  (1)

Where $y_{n={}^{1}/_{-1}}$, indicates the class that point $x_n$ belongs to, n is the number of samples and $x_n$ denotes the real vector of p-dimensional. A scaling is necessary to protect against attributes or variables with higher variance. Using the separating hyperplane, this training data can be viewed that takes:

$w.x + b = o$  (2)

Where b denotes bais, x stands for the input vector and w is the weight vector. The w vector indicates perpendicular to the dividing hyperplane. We can increase the margin by adding the offset variable b. In the absence of b, the hyperplane is compelled to go through the origin, which limits the solution (Luo & Paal, 2022). Because we concern in the largest distance, so we concern about SVM and the


DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والإجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

parallel hyperplanes. The following equations are used to describe the parallel hyperplanes:

$w.x + b = 1$ (3a)

$w.x + b = -1$ (3b)

These hyperplanes can be chosen without any points between them and then attempt to increase their margin when the training data are linear and separated. using geometry, 2/|w| is used to calculate the space between the hyperplane. Thus, we should reduce |w|. To entice data points, we must assure that all i either

$y_i(w.x_i - b) \geq 1 \;\; , 1 \leq i \leq n$ (4)

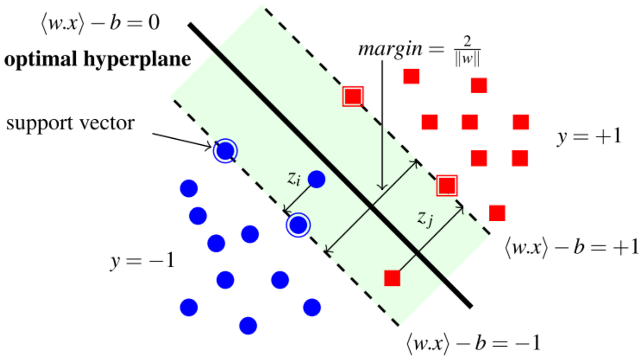


Fig. 2. Two classes (red versus blue) were classified using SVM model

The above figure depicts the idea of SVM, Optimal Canonical Hyperplane (OCH) is a canonical hyperplane with the greatest possible margin. Support Vectors (SVs) are represented as red square and blue circles. Any point $x_i$ that is on the incorrect side of its supporting plane is deemed an error, represented by $z_i$.

Margin $= 2/||w||$ defines the separating hyperplane with the biggest margin which identifies the support vectors closest to it and this satisfies $y_j[w^T.x_j + b] = 1, j = 1.$ OCH would meet the following conditions for all data of the training set.

$y_i[w^T.x_i + b] \geq 1, \;\; i = 1 \ldots \ldots l$ (5)

Where l denotes the amount of data in the training set. To discover the optimal hyperplane with greatest margin, learning machine must reduce $||w^2||$ according to inequality restrictions (equation (5)). The saddle point of the Lagrange's function solves the optimization issue.

$$L_P = L(w, b, \alpha) = \frac{1}{2}\, w^T w - \sum_{i=1}^{l} \alpha_i \,\{y_i[w^T x_i + b] - 1 \qquad (6)$$

Where $\alpha_i$ denotes Lagrange multiplier. The search for optimum saddle points $(w_0, b_0\,, \alpha_0)$ is required because Lagranges should be reduced in terms of w and b and maximized in terms of nonnegative $\alpha_i$ $(\alpha_i \geq 0)$. This issue can be figured out either a dual form (the form $\alpha_i$) or a primal form (the form w & b). The second technique yields interesting findings, and the answer will be discussed in a dual space below. To do this, we employ Karush-Kuhn-Tucker (KKT) criteria for the optimum of a restricted function. Equations (5) and (6) are convex and KKT criteria, respectively,

 DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

that are required and adequate for a maximum of equation (5). Partial differentiation of equation (6) in terms of saddle points $(w_0, b_0, \alpha_0)$

$\frac{\partial L}{\partial w_0} = 0$, i.e, $w_{0=} \sum_{i=1}^{l} a_i y_i x_i$ (7)

$\frac{\partial L}{\partial b_0} = 0$, i.e, $\sum_{i=1}^{l} a_i y_i = 0$ (8)

Replacing equations (7) and (8) in equation (6). We transform from the primal form to the dual form.

$L_d(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i\, y_j \alpha_i \alpha_j x_i^T x_j$ (9)

to identify the optimal hyperplane, a dual Largangian $(L_d)$ should be increased in terms of nonnegative $a_i$ (i.e. $a_i$ should be in the nonnegative quadrant $(L_d)$ and the equality criteria as shown below:

$a_i \geq 0,$ $i = 1$ (10a)

$\sum_{i=1}^{l} \alpha_i\, y_i = o$ (10b)

It should be noted that the dual Lagrangian is stated in respect of training data and is only dependent on the scalar products of input patterns $\left(x_i^T\, x_j\right)$ (Kecman, 2005).

D. Kernel Function of SVM

The function $\Phi$ determines the training vectors $x_i$ to an infinite (possibly, higher) dimensional space. Thereafter, SVM tries to find a linear separating hyperplane accompanied by the largest space in this infinite dimensional space. $C > 0$ is the penalty parameter of the error term. The function of kernel is $K(x_i, x_j) \equiv \Phi(x_i)^T(x_j)$ (Amarappa & Sathyanarayana, 2014). Because SVM has numerous kernel functions, thus choosing the best kernel function is a study topic. Among other things, the choice of kernel function can have a significant impact on the SVM model performance (Panja et al., 2018). Nevertheless, there is no method to determine if the kernel will perform best for a particular pattern recognition issue. Trial is the only method to find the optimal kernel. We may begin with a basic SVM, then the various kernel functions can experiment. Based on the type of the issue, it might be that one kernel performs better than others (Savas & Dovis, 2019).


DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

E. Developed SVM Model

Fig. 3. shows the proposed methodology of oral reading fluency evaluation in SVM.

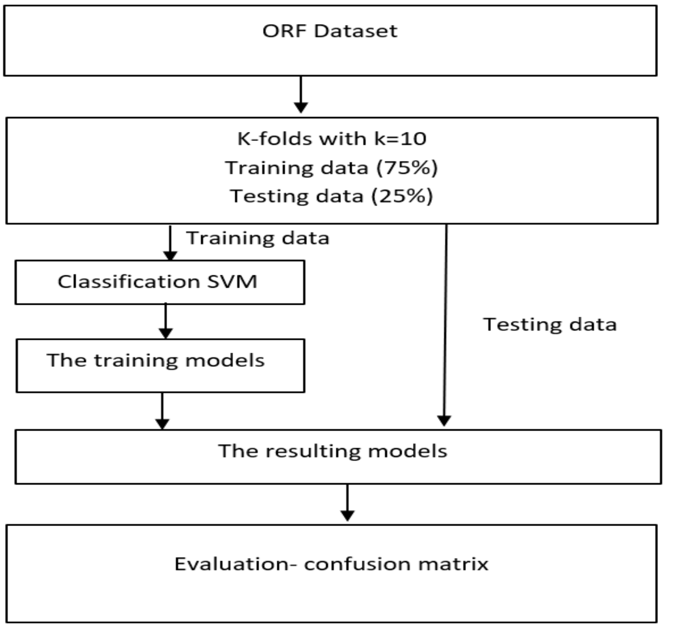


Fig. 3. Flow chart depicting the classification

In this study, the inputs of SVM are expression and volume, phrasing, smoothness and pacing and the output of SVM is the ORF. The ORF data were obtained from the English language department, College of Basic Education, University of Duhok, a total of 100 participants. The data were split into two groups: training data and testing data. The SVM training used the data of training group, while the data of testing group were used to evaluate the model performance. Using K-fold (k=10), the training data used 75% of the data and the rest of 25% data were used as testing data. The ORF output is measured using MDFS scale, see Fig. 1. Matlab 2021 was used as a modeling tool for this study. To generate models, we used SVM with six kernels for training. The models that were created were tested and we used a confusion matrix for evaluation.

F. Performance Measure

The confusion matrix is a popular tool for evaluating classification performance [confusion matrix. The confusion matrix is used to evaluate binary and multi-class classifications as shown in Fig. 4:

| | | Predicted Class | |
|---|---|---|---|
| | | Positive | Negative |
| Actual Class | Positive | TP | FN |
| | Negative | FP | TN |

(a)

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| | | $C_1$ | $C_2$ | … | $C_N$ |
| Actual Class | $C_1$ | $C_{1,1}$ | FP | … | $C_{1,N}$ |
| | $C_2$ | FN | TP | … | FN |
| | … | … | … | … | … |
| | $C_N$ | $C_{N,1}$ | FP | … | $C_{N,N}$ |

(b)

Fig. 4. the examples of confusion matrix. (a) Binary classification. (b) Multi-class classification


DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

There is a group of performance measures that are used for a classification issues' confusion matrix to evaluate an algorithm or contrast the performance of various methods. To assess the efficacy of machine learning algorithms, we utilize several measures, including accuracy, precision, recall, and F1-score. The accuracy is defined as the proportion of properly classified read fluency data to the total number of data, see the below equation:

$$\text{acuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (11)$$

Where true positive (TP) is the number of read fluency accurately categorized into the class they are the property of and the true negative (TN) is the number of read fluency accurately categorized as not belonging to a class. Moreover, the false positive (FP) is the number of read fluency inaccurately categorized to a class, and the false negative (FN) is the number of read fluency inaccurately categorized as not corresponding to a class. To test the accuracy and develop the dependability of the outcomes, we employ 10-fold cross-validation average.

In certain circumstances, where a class represents the majority of sample values in the dataset, the value of accuracy score may not accurately reflect the performance of classifier model. in order to solve this issue and provide the performance of the classifier model, we utilize other performance metrics (recall, precision, and F1- score) (Markoulidakis et al., 2021).

The precision is the proportion of accurately predicted positive read fluency classes to the total number of positive classification predictions, see the below equation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (12)$$

The recall is the proportion of real positive, accurately estimated read fluency classes. As shown in equation (13):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (13)$$

F1-score assesses the medium of precision and recall (Haryanto, A.W. and Mawardi, 2018). See the below equation:

$$\text{F1} - \text{score} = 2 \times \text{Precision} \times \text{Recall} / \text{Precision} + \text{Recall} \qquad (14)$$

## III. Results and Discussion

This part covers the assessment outcomes generated by the algorithm of machine learning for classification in respect of performance metrics and the effect of the kernel on accuracy. In this study, support vector machine (SVM) using Matlab 2021 was utilized in order to predict oral reading fluency (ORF) for the first-year university students of the English language department, College of Basic Education, University of Duhok, Duhok governorate, Iraq. The participants' assessment of ORF was done in 2020 during covid-19 and a total of 100 participants are used for this study. Participants' oral reading fluency was measured using the MDFS scale four dimensions (expression and volume, phrasing, smoothness, and pacing). The four dimensions were used as input parameters and the output was ORF levels. In order to assess the performance of SVM, the dataset was split into two groups (75% for calculating training and 25% for calculating testing). Four input variables, expression and volume, phrasing, smoothness and pacing, were used to evaluate the oral reading fluency.

 DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

The SVM is a supervised learning algorithm that represents data samples in a high-dimensional space. This space specifies a hyperplane to split the datasets and increase the distance between the classes. Because Support Vector Machine is a kernel-sensitive algorithm, selecting kernels for a specific dataset is a sophisticated and difficult option for Data Miner Analysts. We used different kernels to increase accuracy such as Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM and Coarse Gaussian SVM. Using various kernel variables, we can estimate the generalized accuracy. as seen in table 1, each kernel has a different accuracy number. However, it can be seen that SVM with the use of linear kernel provide the best classification accuracy reached 96.2%, see Fig. 5. Thus, it is able to better represent the data classification.

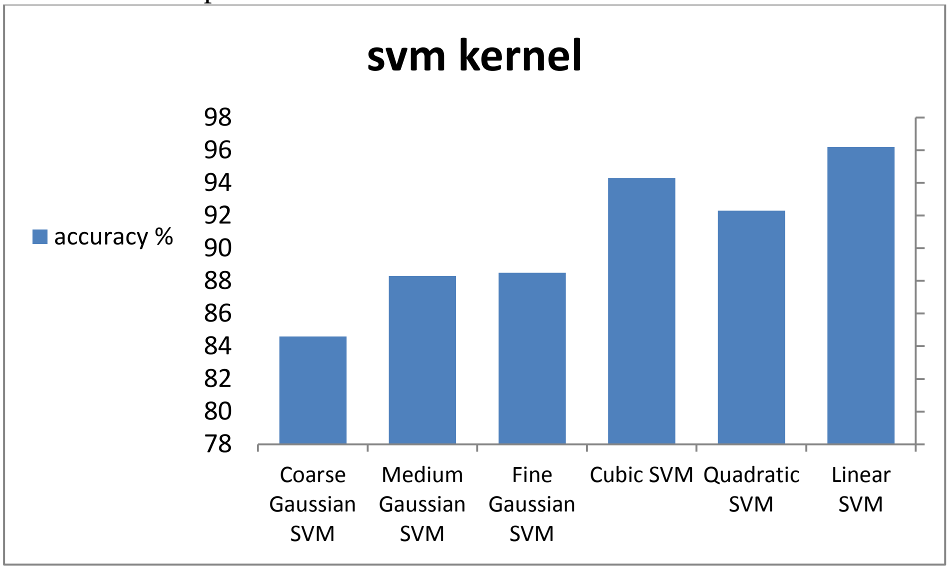


Fig. 5. The accuracy of SVM kernels

The below table shows the exact accuracy for each kernel.

TABLE 1. the exact accuracy number for each SVM kernel

| Kernel type | Accuracy |
|---|---|
| Linear SVM | 96.2 |
| Quadratic SVM | 92.3 |
| Cubic SVM | 94.3 |
| Fine Gaussian SVM | 88.5 |
| Medium Gaussian SVM | 88.3 |
| Coarse Gaussian SVM | 84.6 |

Table (2) presents three evaluation metrics that show the confidentiality of SVM model, these are the precision, recall and F1 score which are calculated using equations (12), (13) and (14).

 DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

TABLE 2. Precision, recall, and F1-score pers for SVM classification.

| | accuracy |
|---|---|
| Precision | 96.1 |
| Recall | 96.1 |
| F1-score | 96.1 |

The results in the above table showed that the values of precision, recall and F1-score for the best SVM linear kernel are 96.1% for all performance metrics. The results of performance metrics for the linear kernel are the best amongst other kernels.

## IV. CONCLUSION

In this study, the ORF has been predicted using SVM, which is one of the algorithms of machine learning. Four parameters have used which are expression and volume, phrasing, smoothness and pacing to predict the ORF for the year 2020 during covid-19. To train SVM, the dataset was divided into 75% for training and 25% for testing. The ORF has been measured using MDFS scale by two raters. The impact of various kernel functions on the ORF predicting performance of SVM model has been investigated. The kernel scale value with the lowest error has been established. The best accuracy value was 96.2% on the use of SVM linear kernel and the performance metrics (precision, recall and F1-score) also recorded the highest accuracy results with the SVM linear kernel reached 96.1%. Based on the results, it is shown that the proposed prediction model is more efficient using the linear SVM kernel. In summary, the approach of SVM is successfully accurate in predicting oral reading fluency from existing data.

## APPENDIX

The text used in this paper is:

**To Marry or Not to Marry!**

Attitudes toward marriage are changing in Japan. In the past, most women were expected to be married by the time they were 25. Women who were not married by then were often thought to have missed out. These single women were sometimes even compared to Christmas cake on December 26th- old and not wanted. Now, things are very different.

According to an opinion poll in a Japanese newspaper The Daily Yomiuri, 73 percent of single Japanese women say they are happy to be single. This is an increase of 10 percent since 2003.

Dr. Sumi Kitade is one of them. She laughs and smiles as she discusses her future. “I will continue my career as a professor,” Dr. Kitade says. “I love my work. It’s very exciting.” When asked about getting married, the 30-year-old professor becomes serious. “To tell the truth,” she replies, “I don’t think I will ever get married. I am happy. I have a wonderful job and I have many friends.”

This attitude is reflected in the increase in the number of single women. In 1970, 18 percent of Japanese women between the ages of 25 and 29 were not married. Thirty years later, that figure had risen to well over 50 percent.

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

The reason can be found in the workplace and at home. On the one hand, more women have full-time jobs than 30 years ago. On the other hand, the traditional role of the wife has not changed. Women are still expected to raise the children and look after the house. They often don't get much help from their husbands. Many women manage the house in addition to doing a full-time job. One survey found that working women spend two hours each day on housework, while men spend about ten minutes. Young working women may choose not to take on these extra responsibilities. Sumi Kitade appears to support this view. "I don't want to quit my job to become someone's slave," says Dr. Kitade.

In the past, women looked for husbands who could offer financial support. Now, women are looking for something different. Kaora Abe, one of Dr. Kitade's colleagues, is 34. She is looking for a husband who will look after the children and share the housework equally. She says she has nothing against marriage. She just hasn't found the right man yet. If her ideal man is out there, she will be very happy. And if he isn't, she says, she'll be happy on her own.

**References**

1. Amarappa, S. and Sathyanarayana, S.V. (2014). Data classification using Support vector Machine (SVM), a simplified approach. Int. J. Electron. Comput. Sci. Eng, 3: 435-445.

2. Ardington, C., Wills, G., Pretorius, E., Mohohlwane, N. and Menendez, A. (2021). Benchmarking oral reading fluency in the early grades in Nguni languages. International Journal of Educational Development, 84, p.102433.
3. Bailly, G., Godde, E., Piat-Marchand, A.L. and Bosse, M.L. (2022). Automatic assessment of oral readings of young pupils. Speech Communication, 138: 67-79.

4. Beck, J.E., Jia, P.E.N.G. and Mostow, J.A.C.K. (2004). Automatically assessing oral reading fluency in a computer tutor that listens. Technology Instruction Cognition and Learning, 2: 61-82.

5. Bolanos, D., Cole, R.A., Ward, W.H., Tindal, G.A., Schwanenflugel, P.J. and Kuhn, M.R. (2013). Automatic assessment of expressive oral reading. Speech Communication, 55(2): 221-236.

6. Cheng, J. (2018). Real-Time Scoring of an Oral Reading Assessment on Mobile Devices. Grantee Submission.

7. Day, R. R. & Yamanaka, J. (2007). Cover to cover1 reading comprehension and fluency. Oxford University Press.
8. Grabe, W. (2010). Fluency in reading—thirty-five years later. Reading in a Foreign Language, 22 (1): 71-83.
9. Haryanto, A.W. and Mawardi, E.K. (2018), September. Influence of word normalization and chi-squared feature selection on support vector machine (svm) text classification. In 2018 International Seminar on Application for Technology of Information and Communication (pp. 229-233). IEEE.

 DOI: https://doi.org/10.33193/IJoHSS.39.2022.492

المجلة الدولية للعلوم الإنسانية والاجتماعية
International Journal on Humanities and Social Sciences
website:www.ijohss.com
Email:editor@ijohss.com
ISSN: 2415 – 4822
العدد (39) نوفمبر 2022
Volume (39) November 2022
IJOHSS

10. Ja'afar, H., Othman, W.M., Kesevan, H.V. and Budi, M.S. (2021). Enhancing Oral Reading Fluency (ORF) Through Computer Assisted Repeated Reading (CARR). Asian Journal of University Education, 17(1): 207-221.

11. Kecman, V. (2005). Support vector machines–an introduction. In Support vector machines: theory and applications (pp. 1-47). Springer, Berlin, Heidelberg.

12. Luo, H. and Paal, S.G. (2022). A data-free, support vector machine-based physics-driven estimator for dynamic response computation. Computer-Aided Civil and Infrastructure Engineering.

13. Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A. and Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. Technologies, 9(4), p.81.

14. Morrison, T., G. & Wilcox, B. (2020). Assessing expressive oral reading fluency. Education Sciences, 10 (59): 1-13.

15. Nation, P. (2014). Developing fluency. In T. Muller, J. Adamson, P. Brown, S. Herder. Exploring EFL Fluency in Asia. (pp. 11-25). Palgrave Macmillan.

16. Panja, S., Chatterjee, A. and Yasmin, G. (2018). July. Kernel Functions of SVM: A Comparison and Optimal Solution. In International Conference on Advanced Informatics for Computing Research (pp. 88-97). Springer, Singapore.

17. Panup, W., Ratipapongton, W. and Wangkeeree, R. (2022). A novel twin support vector machine with generalized pinball loss function for pattern classification. Symmetry, 14(2), p.289.

18. Pisner, D.A. and Schnyer, D.M. (2020). Support vector machine. In Machine learning (pp.101-121). Academic Press.

19. Rasinski, T. V. (2004). Assessing reading fluency. Pacific Resources for Education and Learning (PREL).

20. Samuels, J. (2012). Reading fluency: Its past, present and future. In T. Rasinski, C. Blachowicz, &K. Lems (Eds.), Fluency instruction: Research-based best practices (pp.7-20). New York, NY: The Guilford Press.

21. Savas, C. and Dovis, F. (2019). The impact of different kernel functions on the performance of scintillation detection based on support vector machines. Sensors, 19(23), p.5219.

22. Sen, J., Mehtab, S. and Engelbrecht, A. (2021). Machine Learning: Algorithms, Models and Applications. BoD–Books on Demand.

23. Varol, H.A., Mani, S., Compton, D.L., Fuchs, L.S. and Fuchs, D. (2009). Early prediction of reading disability using machine learning. In AMIA annual symposium proceedings (Vol. 2009, p. 667). American Medical Informatics Association.

 **DOI: https://doi.org/10.33193/IJoHSS.39.2022.492**