Multimedia Tools and Applications (2024) 83:52819–52840
https://doi.org/10.1007/s11042-023-17535-8

Automated reading level classification model based on improved orbital pattern

Rusul Qasim Abed[1] · Melih Dikmen[2] · Emrah Aydemir[3] · Prabal Datta Barua[4] · Sengul Dogan[5] · Turker Tuncer[5] · Elizabeth Emma Palmer[6,7] · Edward J. Ciaccio[8] · U. Rajendra Acharya[9]

Received: 2 August 2022 / Revised: 23 May 2023 / Accepted: 15 October 2023 /
Published online: 14 November 2023

Abstract
Automatic reading level for detection and classification is a challenging problem in machine learning. A multilevel feature extraction-based self-organized model may be useful to overcome this hurdle without using deep learning, which requires an ultra-large sample size. In this work, a novel speech dataset was collected from 57 primary school students by reading a fixed paragraph, and experts labeled these speeches as good, moderate, or bad. We then developed a handcrafted, self-organized learning model. We constructed a novel method using a multilevel feature extraction method, termed improved orbital pattern (IOP) and wavelet packet decomposition (WPD). The proposed IOP generates textural features from the speeches and the used wavelet bands. These extracted features are input to neighborhood components analysis (NCA) to reduce feature dimension. Then the feature set is input to the support vector machine (SVM) classifier to obtain loss values. The output of ten feature vectors of the NCA and SVM classifiers are merged to provide the final feature vector. The most significant 512 features were selected using the NCA feature selection function. These 512 features are classified via the SVM classifier with tenfold cross-validation (CV) and leave-one-subject-out (LOSO) validation strategies. The proposed IOP and WPD-based model yielded an accuracy of 92.75% with a tenfold CV and a 76.18% accuracy using LOSO validation strategies in classifying bad, intermediate, and good reading levels. Our developed model is ready to be validated with more data before its actual usage in schools to aid the teachers.

Keywords Human–computer interface · Teaching/learning strategies · 21st-century abilities · Data science applications in education

1 Introduction

In today's age of rapid information dissemination, effective reading and comprehension skills are vital to fully participate in society and the workplace [1, 2]. Reading is a critical multidimensional skill that should be acquired early [3]. Reading includes linguistic and cognitive processes to facilitate understanding of the meaning of written text [4]. One

Extended author information available on the last page of the article important focus of reading education is fluency, the ability to read a text quickly and accurately. Students who cannot read texts fluently are more hesitant, and reading can become boring and tiring, slowing comprehension skills [5, 6]. Difficulties in fluent and accurate reading, if not appropriately identified and addressed as early as possible, can result in difficulties in student comprehension, knowledge acquisition, and flow, with the possibility of negative impacts on self-esteem and social skills, along with life-long repercussions [7].

We present here a summary of reading skill studies that have been performed in Turkey and suggest how computerized systems, for example, machine learning or data mining, could have improved the speed and depth of study.

1.1 Motivation and method

Machine learning (ML) models are very popular nowadays, and ML models have generally been used in various research areas to simplify tasks [8–11]. Education is one of the disciplines in which machine learning is used most frequently [12, 13]. Our primary aim is to help students with learning difficulties. In the first phase, a detection model has been proposed. This detection model concerns reading level classification. Classification of reading level is of critical importance in primary school education. Early detection of reading difficulties and disorders is critical to rapidly implementing remedial measures tailored to the particular areas of reading that children find challenging. There is potential for artificial intelligence (AI) to assist in the accurate detection of reading difficulties. The main motivation of this study is to propose a novel automated reading level classification model using handcrafted features-based effective learning. We collected a new speech dataset from primary school students who read a standard paragraph. Feature engineering is a crucial study area used to develop machine learning models. We developed a local feature generator called the 'improved orbital pattern' (IOP) in this work. IOP created a feature vector of 2048 in length.

The main motivation of IOP is to extract hidden patterns by analyzing the relationships between multiple orbitals. In contrast, local binary pattern (LBP) [14, 15] only considers a single orbital to extract features. Therefore, in this study, we utilized three orbitals in IOP to extract distinctive multi-orbital patterns similar to Saturn and analyzed the relationships between the orbitals in depth. Briefly, the introduced IOP is considered as the first version of the Saturn patterns, and we have investigated the feature extraction ability of this model in this work. Moreover, the wavelet packet decomposition (WPD) [16] model is used as a subband generator that helps generate high and low-level features. The proposed feature engineering model is a self-organized learning model that selects the most valuable bands according to classification performance. The performances of the bands have been calculated using the SVM classifier [17]. The final feature vector is created by merging the top 10 feature vectors, and the best 512 (per the test results, we attained the best classification performance with 512 features) features are chosen using neighborhood components analysis (NCA) as the feature selector [18]. We employed ten-fold cross-validation (CV) and leave one subject out (LOSO) validation techniques to develop our model.

Briefly, our main motivations are:

- Propose an automated reading level classification model using feature engineering,
- Collect a new speech signal dataset to contribute to speech classification,
- Present a smart assistant model for education.

1.2 Main contributions and novelties

The main contributions and novelties of our IOP-WPD and NCA-based speech classification model are:

- To the best of our knowledge, the presented model is the first self-organized feature engineering model to classify reading levels of primary school students (Bad, Intermediate/Moderate, and Good). In this respect, we are the first research team to propose a speech-based reading-level classification model.
- A new speech dataset was collected from 57 primary school Turkish students. To the best of our knowledge, this topic has no public dataset. Therefore, we collected speech signals from a group of 57 participants.
- A new textural feature extraction-based speech descriptor (IOP) is introduced. We have presented a new kernel and a new paradigm to obtain features.
- A self-organized feature generation model is presented.
- We have used a three-level feature selection methodology.
- To illustrate the robustness of our presented self-organized learning method, tenfold cross-validation, and LOSO cross-validation are employed.
- The presented model attained satisfactory classification performance.

2 Literature review

A few previously presented automated reading level detection are also summarized as follows. Bolanos et al. [19] analyzed reading fluency level. They used 783 one-minute recording corpora for 313 students in the US state of Colorado. They obtained an accuracy rate of 83.30%. Kodan [20] presented a model to evaluate reading difficulties. Data from second, third, and fourth-grade primary school students of Turkey were used. The average number of mistakes in the text was found to be 8.75, 22.16, and 52.78 for independent, teaching, and anxiety levels, respectively. Moreover, the overall duration taken to read the text was 3.73, 4.38, and 6.42 min for independent, teaching, and anxiety, respectively. Xiao and Hu [21] presented a study for English as a second language level assessment using machine learning techniques in Canada. 203 learners from the PIRLS 2016 dataset were analyzed. They attained an accuracy rate of 79.30%. Babayigit [22] investigated the effect of word length on reading ability in Turkish. Data were collected from 342 primary school students. Data were analyzed using descriptive statistical methods (frequency, percentage, mean), and a one-way analysis of variance was performed for samples unrelated to inferential statistics. The study found that first-grade students had an average reading time of 72 ms per word, and second-grade students, third-grade students, and fourth-grade students had an average of 48 ms per word, 28 ms, and 26 ms, respectively. His study reported a significant relationship between students' grade levels and their reading speed according to word length. Ulu [23] evaluated 279 primary school students' reading levels. Problem-solving success was classified as high or low (the two-class problem was defined), depending on their reading speed, reading accuracy percentage, prosodic reading, literal comprehension, and inferential comprehension data. He used discriminant analysis on their collected data. The study determined that fluent reading skills did not affect students' problem-solving abilities. Feature engineering models are faster than deep learning models. Hoskins et al. [24] proposed an approach to detect reading deficits. They evaluated 722 school pupils' data for Carolina automated reading evaluation and dynamic indicators of basic early literacy skills tests. Their study confirmed that the Carolina automated reading evaluation could be used as a reading screening tool. Furthermore, an accuracy rate of 68.84% for binary classification was obtained.

As can be seen from the literature review, the gaps in the previously presented models are:

- The used datasets only detected bad or good. They used only two datasets.
- These datasets were not collected from a homogenous group. This heterogeneous state affected classification accuracy.
- Some models only utilized statistical or manual analysis.

We have presented a new IOP and NCA-based reading level classification model using speeches to fill these gaps.

3 Dataset

We created a dataset involving 57 students from Diyarbakır Kokulupinar primary school in Turkey. The Social and Human Sciences Research Ethics Committee, Firat University, Turkey approved the study. The classroom teacher, who voluntarily supported the study, identified fifty-eight students they considered to have good, average, and poor reading skills at their school. The parents of these students were contacted about the study by the study investigators via telephone to explain the study and obtain permission. The voluntary nature of the study was explained. All parents agreed to their child's participation, although one parent subsequently asked to withdraw his child from the study. The study included 28 female and 29 male students, whose ages varied between 8 and 11 (the average age is 10.15). The reading levels and genders of the students are given in Table 1.

The study population has been tabulated in Table 1, and a fixed/standard paragraph was used to detect the reading level of the participants. A standard reading text is given in Appendix 1. Moreover, the teachers of these students made labeling and we validated these labels according to their reading points.

We called the participants' parents and obtained permission from them for the participation of their child to this study. The collected speech files were split into separate files for each sentence. While naming the resulting files, sequential and underscore character distinction was used in the form of reading level, person ID, and file ID.

The audio files were listened by three different teachers independently, blinded to each child's name. Each sentence was labeled according to the reading level. Some students read some sentences well and others poorly. For this reason, the numbers in each class differed according to the reading level. The number of observations is shown in Table 2. This dataset was uploaded to Kaggle [25].

Table 1 Details of the collected dataset

Reading Level	Female	Male	Total
Bad	3	15	18
Good	14	5	19
Intermediate	11	9	20
Grand Total	28	29	57

Table 2 Number of observations by reading levels

Reading Level	Observation
Bad	402
Intermediate/Moderate	460
Good	437
Total	1299

4 IOP-WPD and NCA based learning model

Our main objective is to propose a new machine learning model for reading-level classification. Thus, we have proposed a new feature engineering. Feature engineering is a vital research area which involves various learning methods, and models in the literature. The most important attribute of the feature engineering models is low time complexity. The main purpose of this work is to present a self-organized feature engineering model. Therefore, a new LBP-like speech descriptor was suggested, called IOP. IOP generates textural features like LBP but uses a special pattern [14, 15]. By using WPD and the suggested IOP, a handcrafted, multilevel feature extraction method was followed. We wanted to depict the learning steps of our proposal. A new approximation concerning the LBP-like feature extraction function has been presented by proposing an IOP descriptor. The presented IOP-WPD and NCA-based model contain IOP and WPD-based multilevel feature vector creation, the feature is chosen to employ NCA, and classification is done by deploying SVM. A summarization of the presented model is shown in schematic form in Fig. 1.

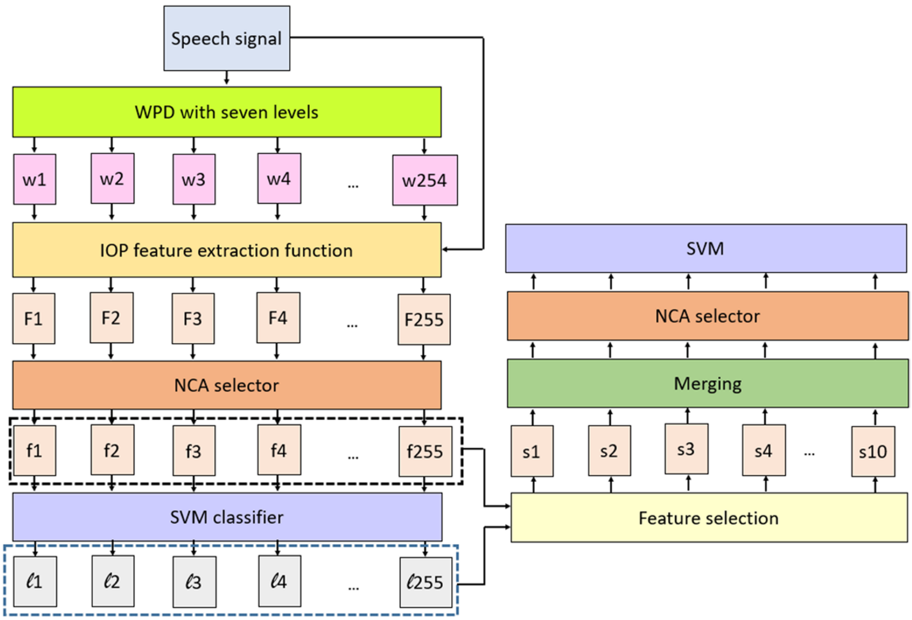

Fig. 1 Schematic illustration of the proposed IOP-WPD and NCA-based model. Here, w: wavelet bands, F: feature vectors, f: selected feature vectors, *l*: loss values and s: selected final feature vectors

Figure 1 indicates that WPD with seven levels has been applied to each speech signal, and 254 wavelet bands have been created, which are each signified with a 'w' in Fig. 1. The presented IOP has been utilized as the main feature extraction function and it generates 255 feature vectors from the used 255 (254 wavelet bands + 1 raw speech signal). The generated feature vectors are shown in Fig. 1 using F. The top 256 features out of 2048 features are selected using NCA, and decreased features (depicted using f) are calculated with the SVM classifier applied to calculate loss values (l1, l2, …, l255). Utilizing loss values, the top 10 feature vectors are chosen (s is utilized to show top features), and these features are merged. Finally, NCA selects the top 512 features, and SVM generates validation predictions.

Moreover, the pseudocode of this model is provided in Algorithm 1.

Input: Speech signal (s)

Output: Prediction vectors (p)

Load the collected speech dataset.
for k=1 to NS **do**
    Read each speech signal
    Apply WPD with seven levels to speech signal and create 254 bands
    $F_1(k, 1{:}2048) = IOP(s);$
    for i=1 to 254 do
        $F_{i+1}(k, 1{:}2048) = IOP(w_i);$
    end for i
    Assign label of the used speech signal.
end for k
for i=1 to 255 **do**
    $idx = fsNCA(F_i, y);$
    for j=1 to 256 **do**
        $f_i(:, j) = F_i(:, idx(j));$
    end for j
    Calculate loss values using SVM and the chosen features.
end for i
Choose the top 10 feature vector using loss values.
Merge the chosen top 10 features.
Choose the most significant 512 features employing $fsNCA$.
Classify these features using SVM with LOSO and 10-fold cross-validation (CV) and calculate p.

Algorithm 1 IOP-WPD and NCA-based speech classification model

In Algorithm 1, NS defines the number of speech in the used dataset, *IOP*(.) represents the IOP feature extraction function, and details of this function are introduced in the feature extraction section, *fsNCA*(., .) is the feature selection with the NCA function, and *idx* denotes sorted indexes. This presented feature engineering model is a parametric learning model. The used parameters have been defined in the Results section.

A detailed explanation of our paradigm is given below, phase by phase. The phases of the presented speech signal classification model are:

- Self-organized feature extraction,
- Feature selection,
- Classification.

Moreover, the steps of our proposal are given below.

4.1 Self-organized feature extraction

The most important phase of the presented IOP- and NCA-based reading level determination model is feature extraction since we have recommended a new function in this work. Moreover, a new multileveled and self-organized feature extraction method has been proposed. This study used a classic machine learning model as the feature extractor. Herein, the implemented main feature creation function was the IOP, which is a novel feature extractor. We were inspired by the local binary pattern (LBP) [26, 27] to propose this model, and a new binary feature extraction approximation has been presented in the presented IOP feature extraction. The steps of our presented self-organized feature extractor are:

Step 1: Apply wavelet packet decomposition (WPD) to the speech signal. In the WPD, symlet four (sym4) has been utilized as the mother wavelet filter. The sym4 filter is one of the most commonly used mother wavelet functions for one-dimensional signals [28, 29]. Using WPD, the wavelet bands have been generated, and these generated frequency bands have been used to create high-level features.

$$w_k = WPD(s, 7), k \in \{1, 2, \dots, 2^8 - 2\} \tag{1}$$

Herein, w_k is the kth wavelet band, and *WPD*() defines wavelet packet decomposition, while s represents the used signal. The parameter and the used input of the WPD are given in Eq. (1). The input is (s) and the parameter is the number of levels. In this work, we have used a seven-levels of WPD. A brief explanation of the WPD is given below.

Wavelet Packet Decomposition (WPD) WPD [28, 29] is a commonly preferred multilevel wavelet decomposition technique. Discrete Wavelet Transform (DWT) generates two bands from a signal and these are named the low-pass filter (L) coefficients band and high-pass filter (H) coefficient band. L bands are the approximate bands, and H bands are the detailed bands. Generally, feature engineering models have used L bands to extract features in the frequency domain. However, H bands help to generate salient features. To solve this problem, WPD has been used for multilevel feature generation. Levels have been generated using both L and H bands with the WPD. Hence, this generates a balanced binary tree. It is a parametric decomposition. The parameters of WPD are the number of levels and the wavelet filter. WPD generates $2^{l+1}-2$ wavelet bands. In this work, l defines the number of levels. In this view, it generates more options than other decomposition models.

Step 2: Extract features from speech and w (bands) deploying IOP. IOP is a textural feature extraction function. We propose an IOP feature extraction function that performs feature extraction on a one-dimensional signal. Specifically, we consider a set of 255 input signals consisting of the raw speech signal and 254 wavelet bands generated by the WPD of the speech signal. We employ the IOP feature extraction function to extract features from each input signal, which applies overlapping blocks with a length of 49 and extracts bits from each block. Using this approach, our method achieves effective feature extraction from the input signals.

$$F_1 = IOP(s) \tag{2}$$

$$F_{k+1} = IOP\left(w_k\right), k \in \{1, 2, \dots, 2^8 - 2\} \tag{3}$$

where F is the extracted features deploying IOP. The steps of IOP are:

Step 2.1: Divide the used one-dimensional input used into overlapping windows with a length of 49.

$$b^i(j) = I(i + j - 1), i \in \{1, 2, \dots, \mathcal{L} - 48\}, j \in \{1, 2, \dots, 49\} \tag{4}$$

Herein, b^i defines the ith overlapping window, with a size/length of 49, I represents the used input signal, and $\mathcal{L}$ is length of the used signal.

Step 2.2: Transform the overlapping blocks into 7×7 sized matrices. Vector to matrix transformation is defined below.

$$\begin{array}{l} m^i(k,l) = b^i(l + 7 \times (k-1)), i \in \{1, 2, \dots, \mathcal{L} - 48\} \\ k \in \{1, 2, \dots, 7\}, l \in \{1, 2, \dots, 7\} \end{array} \tag{5}$$

where m^i is the ith matrix, with a size of 7×7 by generating from b^i.

To apply our suggested pattern, the implemented pattern is depicted in Fig. 2.

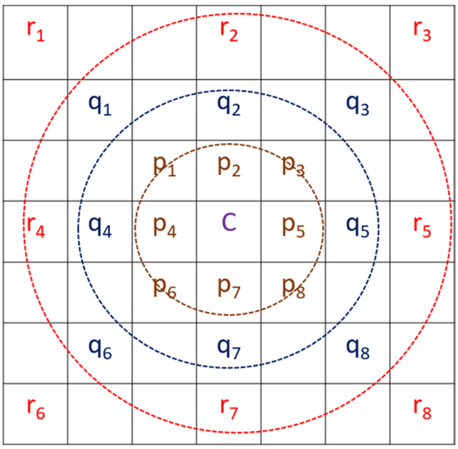

Fig. 2 Pattern of the IOP feature extraction function

In Fig. 2, the employed values are grouped into three categories to create three orbits: r, q, and p. C denotes the center value. By using these values, eight feature map signals have been created. The utilized binary feature extraction function is described below.

The feature extraction function is designed to extract features using three orbits. A minimum 7×7 square matrix is used to extract the features.

Step 2.3: Generate binary features deploying our proposed bit creation approximation.

$$k\left(r_w, q_w, p_w\right) = v_w = \begin{cases} 1, r_w - c < 0 \; and \; q_w - c < 0 \; and \; p_w - c < 0 \\ 2, r_w - c < 0 \; and \; q_w - c < 0 \; and \; p_w - c \geq 0 \\ 3, r_w - c < 0 \; and \; q_w - c \geq 0 \; and \; p_w - c < 0 \\ 4, r_w - c < 0 \; and \; q_w - c \geq 0 \; and \; p_w - c \geq 0 \\ 5, r_w - c \geq 0 \; and \; q_w - c < 0 \; and \; p_w - c < 0 \\ 6, r_w - c \geq 0 \; and \; q_w - c < 0 \; and \; p_w - c \geq 0 \\ 7, r_w - c \geq 0 \; and \; q_w - c \geq 0 \; and \; p_w - c < 0 \\ 8, r_w - c \geq 0 \; and \; q_w - c \geq 0 \; and \; p_w - c \geq 0 \end{cases}, \tag{6}$$

$$w \in \{1, 2, \ldots, 8\}$$

In Eq. (6), the employed kernel is defined, and values (v) range from 1 to 8 are created. We obtain eight bits of resolution with eight categories using these values. Binary features extraction using these values is defined below.

$$bit_w^h = \begin{cases} 0, h \neq k\left(r_w, q_w, p_w\right) \\ 1, h = k\left(r_w, q_w, p_w\right) \end{cases}, h \in \{1, 2, \ldots, 8\} \tag{7}$$

Step 2.4: Generate map signals using the bits extracted.

$$map^h(i) = \sum_{t=1}^{8} bit^h \times 2^{8-h}, , i \in \{1, 2, \ldots, \mathcal{L} - 48\} \tag{8}$$

Step 2.5: Extract histograms of each map signal. The histogram is a counter that shows how many times the values of the map signal are repeated. Using these values, features are obtained. Histogram-based feature extractors are useful for classification as they give the probability of repeating each value of the map signals. Hence, we have proposed a histogram-based feature extractor.

Step 2.6: Merge the eight generated histograms and obtain a feature vector of 2048. We have used the concatenation function in this step.

Steps 2.1–2.6 define the proposed IOP feature extraction/generation function. We used orbits to get features in the presented IOP feature extraction function. Moreover, we have proposed a new kernel to extract valuable features. Three orbits have been utilized. This model is a new LBP-like model and extracts more features than LBP. Therefore, IOP generates more hidden patterns than other LBP-like feature extraction functions.

Step 3: Decrease the dimension of the generated features (F) by deploying the NCA selector. 256 out of 2048 generated features are obtained.

$$idx^h = fsNCA\left(F_h, y\right), h \in \{1, 2, \dots, 255\} \quad (9)$$

$$f_h(k, l) = F_h\left(k, idx^h(l)\right), k \in \{1, 2, \dots, NS\}, l \in \{1, 2, \dots, 256\} \quad (10)$$

Step 4: Compute error values (loss array) utilizing the SVM classifier with a tenfold CV.
Step 5: Choose the top 10 feature vectors (top f vectors) deploying loss values.
Step 6: Merge the chosen top 10 vectors. We have used the concatenation function in this step, like in Step 2.6.

The given six steps above define the presented feature creation model, and 2560 features are generated.

4.2 Feature selection

As is elucidated in Sect. 3.1, NCA is used to choose the best/most valuable features. Thus, NCA has been considered to choose 512 salient features from the 2560 features created [18, 30]. In order to determine the best working feature selector, we tested four feature selection functions (Chi2, ReliefF, mRMR, and NCA) and the best performing feature selector is NCA, and the 512 features provide satisfactory results. NCA is a distance-based learning method similar to the k-nearest neighbor algorithm. Moreover, the stochastic gradient descent (SGD) method is used with NCA to easily select the most useful features. Herein, the fitness function of this solver is an L1-norm-based distance.

Step 7: Choose the most useful 512 features out of 2560 features created.

4.3 Classification

The last phase is classification. Herein, we proposed a feature engineering model. Therefore, we have used a shallow/conventional classifier to show the high classification capability of the selected features. SVM is the most widely used conventional classifier in the literature and has various kernels. Hence, we have chosen SVM as the classifier [17]. The hyperparameters of the SVM used are:

Kernel: polynomial order
Degree: 3rd
Box constraint: 1
Coding: One-Vs-One
Validation: tenfold CV / LOSO

In this work, we have used Cubic SVM, and default settings were engaged to create a validation prediction vector.

Step 8: Classify 512 features using SVM with LOSO/tenfold CV strategies.

5 Results

We have presented a self-organized feature extraction model and a three-leveled feature selection phase to obtain high classification performance in this work since we have addressed the reading level classification. The performance of this work has been provided in this section.

5.1 Model construction

We presented a parametric feature engineering architecture, which was programmed with a simple computer. We only required 8 GB of main memory, a simple CPU (over 3 GHz frequency), and the MATLAB (2021a) programming environment to program the IOP and NCA-based feature engineering model. There was no need to use parallel programming or an expensive graphical processing unit card for this system. Model parameters are given below (Table 3).

5.2 Performance evaluation metrics

Four widely used performance evaluation parameters are employed in this work, given as follows. In papers noted in the published literature, classification has been evaluated by calculating the accuracy of their performance; hence, we used accuracy. However, accuracy is an insufficient performance evaluation parameter for unbalanced datasets. Therefore, recall (classification accuracy of each class), precision (detection rate calculator) and F1-score (harmonic mean of the precision and recall) have been used in this model. The mathematical definitions of the used performance evaluation metrics are given below.

$$a = \frac{tp + tn}{tp + tn + fp + fn} \tag{11}$$

$$r = \frac{tp}{tp + fn} \tag{12}$$

$$pr = \frac{tp}{tp + fp} \tag{13}$$

$$f1 = \frac{2 \times r \times pr}{r + pr} \tag{14}$$

where a is accuracy, r means recall, pr represents precision and $f1$ is F1-score. The number of true positives (tp), true negatives (tn), false negatives (fn) and false positives (fp) are used to compute the performance matrices.

5.3 Classification performance

The classification result of the proposed model is given in this section. In the classification phase, we used two validation techniques, and these are tenfold CV and LOSO CV. For the tenfold CV method, the data is separated into ten folds, and this division

Table 3 Parameters of the presented IOP-based model

Method	Parameters	Output
WPD	Wavelet filter: sym4 Number of levels: 7	254 bands are generated
IOP	49-sized overlapping blocks and a new feature extraction kernel	2048 features are extracted
Feature extraction	IOP is applied to 254 wavelet bands and the raw speech signal	255 feature vectors and the length of each feature vector is 2048
NCA	Solver: SGD Verbose: No convergence summary Tuning iteration: Half of the number of observations Tuning initial rate Number of tuning iterations: 20 Tuning subset size: 100	The most valuable 256 out of 2048 features are selected from each feature vector
Loss calculation	SVM with tenfold CV	255 misclassification rates are calculated
Feature selection	Greedy	The top 10 feature vectors per the calculated misclassification rates
Merging	Concatenation	Feature vector with a length of 2560 is calculated
NCA	Solver: sgd Verbose: No convergence summary	The top 512 features are selected
SVM	Cubic SVM Validations: tenfold CV and LOSO CV	The result is calculated by deploying this classifier

is generally performed randomly. In each fold, classification performance is calculated and the average classification performance is given as a general result. The LOSO CV separates data by subjects. In this validation method (LOSO CV), the first subject has been utilized for testing and the others have been used for training. LOSO CV renders real-world results of the used datasets but is slower than other validation techniques. We obtain the robust results using these validation strategies. The confusion matrices obtained using tenfold CV, and LOSO validation are depicted in Tables 4 and 5.

Table 4 shows the presented IOP-WPD and NCA-based method, which attained a classification accuracy of 92.75%. Moreover, category-wise results are displayed in Table 4. According to Table 4, the most accurate class attained a 96.10% recall rate.

The considered second validation technique is LOSO. The confusion matrix obtained for LOSO validation is shown in Table 5.

Table 5 illustrates that our paradigm achieved a 76.18% accuracy using the LOSO validation technique. In addition, the best recall rate of 80.50% is obtained for the Good class.

Table 4 Confusion matrix using tenfold CV

Actual outputs	Predicted outputs		
	Bad	Moderate	Good
Bad	365	27	10
Moderate	11	419	29
Good	2	15	419
Recall	0.9080	0.9129	0.9610
Precision	0.9656	0.9089	0.9148
F1	0.9359	0.9109	0.9374

Table 5 Confusion matrix using LOSO

Actual outputs	Predicted outputs		
	Bad	Moderate	Good
Bad	321	51	30
Moderate	61	316	82
Good	9	76	351
Recall	79.85	68.85	80.50
Precision	82.10	71.33	75.81
F1	80.96	70.07	78.09

Also, the overall results obtained for tenfold CV and LOSO are tabulated in Table 6.

In addition, the suggested IOP-WPD and NCA-based speech classification model is self-organized. Each feature vector has been evaluated to create the final/optimal feature vector, and individual accuracies calculated are described in Fig. 3.

Figure 3 shows that the best accuracy was 88.36% using w7 (7th wavelet subband). We have chosen the top 10 feature vectors to construct a final feature vector. The ten best feature vectors were: 8th (extracted from w7), 32nd (extracted from w31), 4th (extracted from w3), 16th (extracted from w15), 64th (extracted from w65), 129th (extracted from w128), 65th (extracted from w64), 2nd (extracted from w1), 33rd (extracted from w32) and 1st (extracted from original speech signal). By concatenating these feature vectors and selecting the top 512 by deploying NCA, the highest classification accuracy is improved from 88.36% to 92.75%.

5.4 Time complexity analysis

Herein, we have presented a self-organized feature engineering model. The proposed IOP-NCA-based model uses IOP and WPD in the feature extraction layer. Therefore, the time complexity of this layer is equal to $O(knlogn)$, where k defines the number of signals, n is

Table 6 Overall performance metrics using tenfold CV and LOSO strategies

Results (%)	LOSO	tenfold CV
Accuracy	76.18	92.75
Precision	76.41	92.98
Recall	76.40	92.73
F1	76.37	92.81

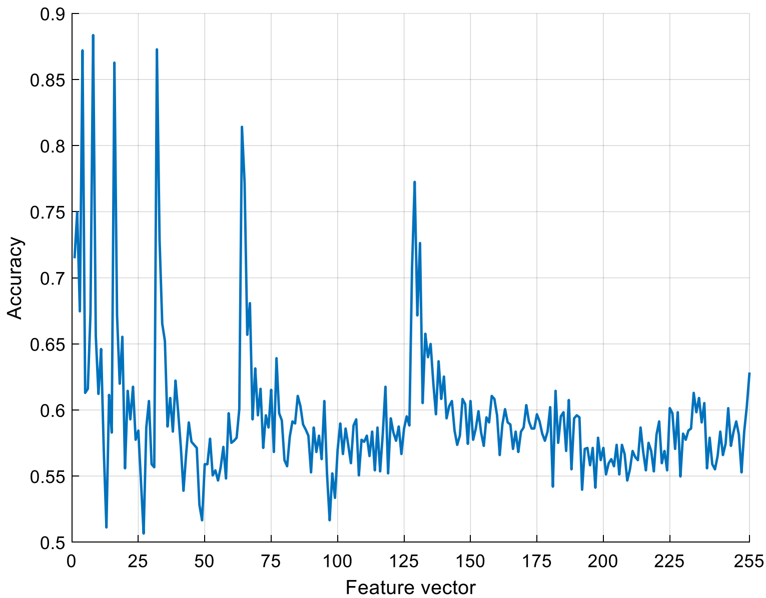

Fig. 3 Graph of accuracy versus features using SVM classifier with tenfold CV strategy

the length of the used signals. The NCA selector was used in the feature selection step. The computational complexity of this phase is $O(kN)$, where N refers to the time complexity coefficient of the NCA selector. To choose the most accurate 10 feature vectors, we have used the SVM classifier and the time burden of this phase is $O(kS)$, and S defines the time complexity coefficient of the SVM classifier. Ten feature vectors are merged and then forwarded to the NCA selector to select the most significant 512 features. Therefore, the burden of this phase is equal to $O(tN)$, where t defines the number of selected features and $t < k$. In the last phase, SVM is used to render the prediction vector, and the computational burden of the SVM is $O(S)$. According to this calculation, the total time burden of this model is $O(knlogn + kN + kS + tN + S) \cong O(knlogn + kN + kS)$. In this mode, this model has a linear time complexity.

6 Discussions

This work introduces a new self-organized feature extraction model using handcrafted features. The presented feature extractor, IOP, generates textural features from speech signals. WPD has been employed to create bands for high-level feature generation. Our proposed system chooses the most informative features using the most significant bands. Hence, we are influenced by deep learning (DL) models. These DL models have various feature extraction levels and can create the most valuable features. Therefore, deep learning networks, especially convolutional neural networks (CNNs), have been implemented in this work. However, they have a time complexity problem, as they need to optimize millions of parameters. Our proposal to address these issues is a handcrafted feature-based model with multilevel feature extraction architecture used to create the most valuable high-level features. Low-level features are extracted from the IOP. Hence, this proposed model mimics CNN. Also, our proposal does not need to set millions of parameters. Therefore, it has lower time complexity as compared with deep learning models. Our model generated 255 feature vectors and has a self-organization ability because it can choose the top feature vector to create the final features. Furthermore, LOSO and tenfold CV have been employed to confirm the robustness of the model.

Our used classifier is compared to conventional classifiers, i.e., the decision tree (DT) [31], linear discriminant (LD) [32], k nearest neighbor (kNN) [33], subspace discriminant (SD) [34], and artificial neural network (ANN) [35]. The obtained comparative results using a tenfold CV are presented in Fig. 4. We have used the same parameters of our model to get these comparatively results. They used the same feature extraction and feature selection parameters. Our main goal is to illustrate the superiority of our used combination.

The second-best classifier is ANN, which attained a 90.21% classification accuracy. SD attained a 90.13% accuracy. The worst classifier for tenfold CV was LD, which reached 76.18% accuracy. Furthermore, LOSO validation is used, and the obtained accuracies are denoted in Fig. 5.

It can be noted from LOSO validation that the used classifiers: DT, LD, SVM, kNN, SD, and ANN attained 64.38%, 61.98%, 76.18%, 68.15%, 72.16%, and 71.62% accuracies, respectively. Furthermore, it can be noted from Figs. 4 and 5 that the best classifier is SVM, and it is chosen for this work. Moreover, we have used three SVM kernels and these are (i) Gaussian, (ii) linear and (iii) polynomial. The test results of these SVMs with a tenfold CV are depicted in Fig. 6.

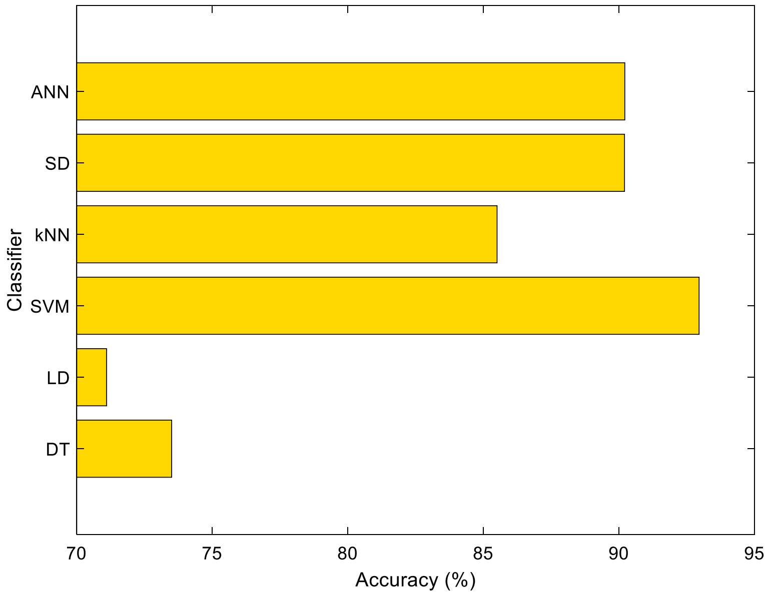

Fig. 4 Accuracies (%) obtained for different classifiers with tenfold CV

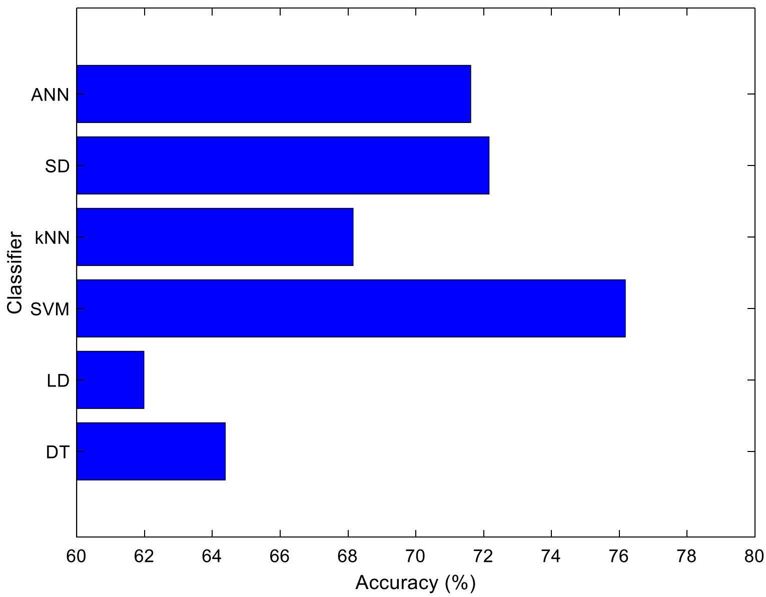

Fig. 5 Accuracies (%) obtained for different classifiers with LOSO validation

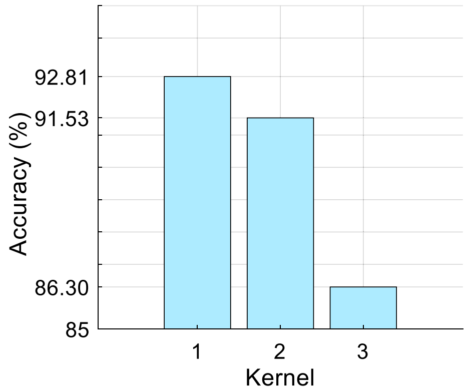

Fig. 6 Classification performance of the SVM with three kernels. 1: Polynomial, 2: Linear, 3: Gaussian

In Fig. 6 it is evident that the most accurate kernel is polynomial. Therefore, we have selected this kernel to obtain classification results.

We have employed four feature selectors to obtain comparative results for our model, namely Chi2 [36], ReliefF [37], mRMR (minimum redundancy maximum relevance) [38], and NCA [18]. We have employed identical quality indices for these feature selectors, and the classification accuracies calculated using them are presented in Fig. 7.

As illustrated in Fig. 7, NCA is found to be the most effective feature selector for our problem.

This research topic is a very new focus area of research. Hence, other publications concerning this topic are not yet available. The closest topic to our model concerns reading fluency detection. The datasets of these works generally contain two classes; hence, they address the binary classification problem.

We highlight the advantages and limitations of this research per the obtained findings:

Advantages:

- To the best of our knowledge, this is the first fully-automated level of reading quality detection work developed for primary school students.
- A novel speech descriptor—IOP, is presented to extract salient low-level features. IOP is a new one-dimensional textural feature extraction function.
- A WPD-based multilevel feature extraction method has been used to extract high-level features. By extracting features from bands of the WPD, features are generated at a high-level.

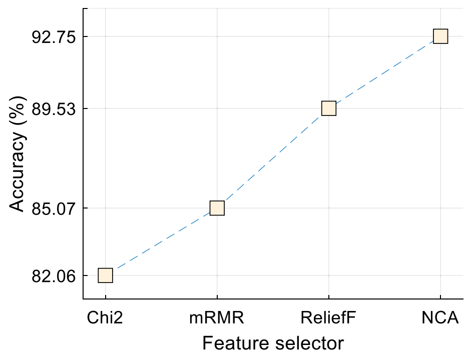

Fig. 7 Accuracies obtained using various feature selectors with a tenfold CV strategy

- A self-organized, completely automated model is proposed. Using this structure, the meaningful bands for feature extraction have been chosen and the redundant ones are eliminated.
- tenfold CV and LOSO validations were employed to obtain robust results.
- The best conventional/shallow classifier was chosen (see Figs. 4 and 5).
- The proposed IOP-WPD and NCA-based model is a handcrafted method. Thus, manual fine-tuning of millions of parameters is not required.

Limitations:

- We have evaluated 57 students' data in this work. In the future, a larger speech dataset will be used to obtain a confirmatory result.
- We used the classifier with default settings. Hyperparameter fine-tuning can be conducted in future works to obtain better performance.

As evident above, we have proposed a highly accurate and lightweight model. We can propose an embedded smart application for future comparisons using our model. In the training phase, we can determine the best bands and feature combinations. Our optimal model only generates features from w_7, w_{31}, w_3, w_{15}, w_{65}, w_{128}, w_{64}, w_1, w_{32} and raw speech signals. Here, w signifies the wavelet bands. Using a look-up table (this table contains the indexes of the most significant features and is generated from training phase), the most important of 512 features have been selected (without redeploying NCA selector) and these features are classified using a pretrained SVM classifier. This model can be applied to other signal classification models. The optimal wavelet bands and the indices of the most significant features have been selected in the training phase. In the testing phase, the wavelet bands have been generated by applying WPD, and the features have been generated by deploying the IOP to address the most informative bands. Other phases (feature selection and classification) are very simple since we have used an array to select the most important features, and these features have been classified using a pretrained SVM.

7 Conclusions

This work presents a novel handcrafted speech classification model to automatically detect primary school students' reading level ability. We have proposed a novel improved orbital pattern (IOP) and wavelet packet decomposition (WPD) techniques. IOP generates textural features from the original speech and frequency band (WPD generates frequency bands). The proposed model is fully automated, as it can select the most salient final features and perform classification. In this work, we have used NCA to choose the 512 most significant features, and these features are classified with the SVM classifier. Our method attained 92.75% and 76.18% accuracies using tenfold CV and LOSO strategies, respectively. A limitation of this work is that we have used only a small number of students' data in each of the three classes. We plan to use more primary school students' data, and develop a more robust model in the future.

In the near future, we will develop a next-generation smart mobile application to detect pupils' reading levels. We plan to collect a larger dataset in various languages to implement this application. Using attention-based explainable models, we can detect the features of the reading speeches, which can assist teachers with their student evaluations [39]. Furthermore, dyslexia can be detected by using this type of application.

Appendix 1

The used stable reading text and translation of this text are given below.

Reading text in Turkish.

“Bir kartal denizden uzak bir dağ yolunun kenarında yuva kurdu. Kartalın orada yavruları oldu. Bir gün pençesinde kocaman bir balıkla yuvasına geldi kartal. Yuva yaptığı ağacın çevresinde çalışan insanlar vardı. Balığı gördüklerinde ağacın etrafında toplanıp bağırmaya, kartala taş atmaya başladılar. Balık sonunda kartalın pençesinden kayıp yere düştü. Adamlar balığı alıp gittiler. Kartal yuvasının bir köşesine çekilip tünedi. Yavruları ise havaya başlarını dikip yiyecek, yiyecek diye bağrışmaya başladılar. Oysa kartal çok yorulmuş denize kadar uçacak gücü kalmamıştı. Yuvasına iyice yerleşip yavrularını kanatlarının altına aldı. Onları sevdi, okşadı ve küçücük tüylerini düzeltti. Sanki “Ne olur birazcık sabredin!” diye yalvarıyordu onlara. Fakat yavrular okşandıkça seslerini daha da yükseltip bağrışmaya devam ettiler. Kartal uçtu ve daha yüksek bir dala kondu. Yavrular, anneleri uçup gidince daha da acıklı bir sesle bağrıştılar. Sonunda kartal çaresizlik içinde acı bir çığlık attı ve kanatlarını açıp ağır ağır denize doğru uçtu. Anne kartal akşam olup geç vakit yuvaya dönerken ağır ağır ve alçaktan uçmaktaydı. Yine pençelerinde kocaman bir balık vardı. Ağaca yaklaşırken çevrede başkaları var mı diye etrafı kolaçan etti bu kez. Güven içinde olduğunu hissettikten sonra kanatlarını kısıp yuvasının bir ucuna kondu. Yavru kartallar gagalarını açıp boyunlarını uzattılar. Anne kartal ise balığı parçaladı ve başladı yavrularını doyurmaya.”

The English translation of the above paragraph is given below.

"An eagle nested by the side of a mountain road far from the sea. The eagle had hatchlings there. One day, an eagle came to its nest with a huge fish in its talons. There were people working around the tree where he made his nest. When they saw the fish, they gathered around the tree and started shouting at the eagle and throwing stones. The fish eventually slipped from the eagle’s talons and fell to the ground. The men took the fish and left. The eagle stepped back and perched on a corner of its nest. The hatchlings, on the other hand, raised their heads and started shouting for food, food. However, the eagle was so tired that it did not have the strength to fly into the sea. She settled in her nest and took her hatchlings under her wing. She loved them, touched them, and straightened their tiny feathers. "Please be patient!" she begged them. But as the hatchlings were attended, they kept raising their voices and shouting. The eagle flew and landed on a higher branch. The hatchlings cried out even more pathetically as their mother flew away. Finally, the eagle let out a cry of despair and spread its wings and flew slowly towards the sea. The mother eagle was flying slowly and low, as it was late in the evening, and was returning to her nest. It also had a huge fish in its paws. As she approached the tree, this time she looked around to see if anyone else was around. Feeling secure, she folded her wings and perched on one end of her nest. The baby eagles opened their beaks and stretched their necks. The mother eagle crushed the fish and began to feed her young."

Funding The authors state that this work has not received any funding.

Declarations

Conflict of interest The authors of this manuscript declare no conflicts of interest.

Institutional Review Board Statement This research has been approved on ethical grounds by the Social and Human Sciences Research Ethics Committee, Firat University, Social and Human Sciences Research Centre Ethics Board on 23 November 2021 (23–5).

References

1. Khumsamart S (2022) Innovative Management Strategies for the Office of Non-Formal and Informal Education in the Digital Age. J Manag Bus Healthcare, Educ 1(2):1–21
2. Ribble M, Park M (2022) The digital citizenship handbook for school leaders: Fostering positive interactions online. Int Soc Technol Educ, Washington, DC.
3. Liu C, Chung KKH, Tang PM (2022) Contributions of orthographic awareness, letter knowledge, and patterning skills to Chinese literacy skills and arithmetic competence. Educ Psychol 42(5):530–548
4. Pae HK (2022) Toward a script relativity hypothesis: Focused research agenda for psycholinguistic experiments in the science of reading. J Cult Cogn Sci 6:1–21
5. Petersen SE, Ostendorf M (2009) A machine learning approach to reading level assessment. Comput Speech Lang 23(1):89–106
6. Ecalle J, Magnan A, Auphan P, Gomes C, Cros L, Suchaut B (2022) Effects of targeted interventions and of specific instructional time on reading ability in French children in grade 1. Eur J Psychol Educ 37(3):605–625
7. Middleton AE, Farris EA, Ring JJ, Odegard TN (2022) Predicting and evaluating treatment response: Evidence toward protracted response patterns for severely impacted students with dyslexia. J Learn Disabil 55(4):272–291
8. Kwon S (2021) Att-Net: Enhanced emotion recognition system using lightweight self-attention module. Appl Soft Comput 102:107101
9. Kwon S (2020) CLSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network. Mathematics 8(12):2133
10. Ahmad F, Ikram S, Ahmad J, Ullah W, Hassan F, Khattak SU, Rehman IU (2020) GASPIDs Versus Non-GASPIDs-Differentiation Based on Machine Learning Approach. Curr Bioinform 15(9):1056–1064
11. Ullah W, Muhammad K, UlHaq I, Ullah A, UllahKhattak S, Sajjad M (2021) Splicing sites prediction of human genome using machine learning techniques. Multimed Tools Appl 80(20):30439–30460
12. Shaikh AA, Kumar A, Jani K, Mitra S, García-Tadeo DA, Devarajan A (2022) The Role of Machine Learning and Artificial Intelligence for making a Digital Classroom and its sustainable Impact on Education during Covid-19. Mater Today: Proceed 56:3211–3215
13. Alam A (2022) Employing Adaptive Learning and Intelligent Tutoring Robots for Virtual Classrooms and Smart Campuses: Reforming Education in the Age of Artificial Intelligence. In: Advanced Computing and Intelligent Technologies. Springer, pp 395–406. https://doi.org/10.1007/978-981-19-2980-9_32
14. Luo Q, Su J, Yang C, Silven O, Liu L (2022) Scale-selective and noise-robust extended local binary pattern for texture classification. Pattern Recogn 132:108901
15. Wei J, Lu G, Yan J, Liu H (2022) Micro-expression recognition using local binary pattern from five intersecting planes. Multimed Tools Appl 81:1–26. https://doi.org/10.1007/s11042-022-12360-x
16. Nayak SK, Jarzębski M, Gramza-Michałowska A, Pal K (2022) Automated Detection of Cannabis-Induced Alteration in Cardiac Autonomic Regulation of the Indian Paddy-Field Workers Using Empirical Mode Decomposition, Discrete Wavelet Transform and Wavelet Packet Decomposition Techniques with HRV Signals. Appl Sci 12(20):10371
17. Chaabane SB, Hijji M, Harrabi R, Seddik H (2022) Face recognition based on statistical features and SVM classifier. Multimed Tools Appl 81(6):8767–8784
18. Goldberger J, Hinton GE, Roweis S, Salakhutdinov RR (2004) Neighbourhood components analysis. Adv Neural Inf Process Syst 17:513–520

19. Bolaños D, Cole RA, Ward WH, Tindal GA, Hasbrouck J, Schwanenflugel PJ (2013) Human and automated assessment of oral reading fluency. J Educ Psychol 105(4):1142
20. Kodan H (2017) Determination of Reading Levels of Primary School Students. Univ J Educ Res 5(11):1962–1969
21. Xiao Y, Hu J (2019) Assessment of optimal pedagogical factors for Canadian ESL learners' reading literacy through artificial intelligence algorithms. Int J English Linguist 9(4):1–14
22. Babayigit O (2018) Evaluation of Reading Rates according to Word Length of Primary School Students. Mehmet Akif Ersoy University Journal of Education Faculty 46:409–427. https://doi.org/10.21764/maeuefd.330831
23. Ulu M (2017) The Effect of Reading Comprehension and Problem Solving Strategies on Classifying Elementary 4th Grade Students with High and Low Problem Solving Success. J Educ Train Stud 5(6):44–63
24. Hoskins WH, Hobbs WI, Eason MJ, Decker S, Tang J (2021) The design and implementation of the Carolina Automated Reading Evaluation for reading deficit screening. Comput Human Behav Rep 4:100123. https://doi.org/10.1016/j.chbr.2021.100123
25. Kaggle (2021) Reading Sounds of According to Reading Levels. http://www.kaggle.com/dataset/9b7cfc0b0ad942b79585629b5c66d17687830c6f5f388bdb4ae5c7dd8a066d23. Accessed 10.05.2022
26. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
27. Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. European conference on computer vision. Springer, pp 469–481
28. Raju G Wavelet transform and projection profiles in handwritten character recognition-A performance analysis. In: 2008 16th International Conference on Advanced Computing and Communications, 2008. IEEE, pp 309–314. https://doi.org/10.1109/ADCOM.2008.4760466
29. Abo-Zahhad M, Rajoub BA (2002) An effective coding technique for the compression of one-dimensional signals using wavelet transforms. Med Eng Phys 24(3):185–199
30. Tuncer T, Ertam F, Dogan S, Subasi A (2020) An automated daily sports activities and gender recognition method based on novel multikernel local diamond pattern using sensor signals. IEEE Trans Instrum Meas 69(12):9441–9448
31. Tariq A, Yan J, Gagnon AS, Riaz Khan M, Mumtaz F (2022) Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. Geo-spatial Inf Sci 1:1–19. https://doi.org/10.1080/10095020.2022.2100287
32. Khan MT, Sha'ameri AZ, Zabidi MMiA, Chia CC (2022) FHSS Signals Classification by Linear Discriminant in a Multi-signal Environment. In: Proceedings of the International e-Conference on Intelligent Systems and Signal Processing, Springer, pp 143–155. https://doi.org/10.1007/978-981-16-2123-9_11
33. Uddin S, Haque I, Lu H, Moni MA, Gide E (2022) Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep 12(1):1–11
34. Sharma SK, Vijayakumar K, Kadam VJ, Williamson S (2022) Breast cancer prediction from microRNA profiling using random subspace ensemble of LDA classifiers via Bayesian optimization. Multimed Tools Appl 81(29):1–21. https://doi.org/10.1007/s11042-021-11653-x
35. Gururaj N, Vinod V, Vijayakumar K (2022) Deep grading of mangoes using convolutional neural network and computer vision. Multimed Tools Appl 82:39525–39550. https://doi.org/10.1007/s11042-021-11616-2
36. Liu H, Setiono R Chi (1995) 2: Feature selection and discretization of numeric attributes. In: Proceedings of 7th IEEE International conference on tools with artificial intelligence. IEEE, pp 388–391. https://doi.org/10.1109/TAI.1995.479783
37. Kononenko I (1994) Estimating attributes: Analysis and extensions of RELIEF. European conference on machine learning. Springer, pp 171–182
38. Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8):1226–1238
39. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR (2022) Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput Methods Programs Biomed 226:107161. https://doi.org/10.1016/j.cmpb.2022.107161

Authors and Affiliations

Rusul Qasim Abed[1] · Melih Dikmen[2] · Emrah Aydemir[3] · Prabal Datta Barua[4] · Sengul Dogan[5] · Turker Tuncer[5] · Elizabeth Emma Palmer[6,7] · Edward J. Ciaccio[8] · U. Rajendra Acharya[9]

✉ Sengul Dogan
sdogan@firat.edu.tr

Rusul Qasim Abed
rusol.kasem1190@gmail.com

Melih Dikmen
mdikmen@firat.edu.tr

Emrah Aydemir
emrahaydemir@sakarya.edu.tr

Prabal Datta Barua
Prabal.Barua@usq.edu.au

Turker Tuncer
turkertuncer@firat.edu.tr

Elizabeth Emma Palmer
elizabeth.palmer@unsw.edu.au

Edward J. Ciaccio
ciaccio@columbia.edu

U. Rajendra Acharya
Rajendra.Acharya@usq.edu.au

1 Department of Computer Engineering, College of Engineering, Kirsehir Ahievran University, Kirsehir, Turkey

2 Faculty of Education, Fırat University, Elazig, Turkey

3 Department of Management Information Systems, Management Faculty, Sakarya University, Sakarya, Turkey

4 School of Business (Information System), University of Southern Queensland, Toowoomba, Australia

5 Department of Digital Forensics Engineering, College of Technology, Firat University, Elazig, Turkey

6 Centre of Clinical Genetics, Sydney Children's Hospitals Network, Randwick 2031, Australia

7 School of Women's and Children's Health, University of New South Wales, Randwick 2031, Australia

8 Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA

9 School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia