Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution

RUBA OBIEDAT[1], **RANEEM QADDOURA**[2], **ALA' M. AL-ZOUBI**[1,3], **LAILA AL-QAISI**[4], **OSAMA HARFOUSHI**[1], **MO'ATH ALREFAI**[1], **AND HOSSAM FARIS**[1,2,5]

[1]King Abdullah II School for Information Technology, The University of Jordan, Amman 11942, Jordan
[2]School of Computing and Informatics, Al Hussein Technical University, Amman 11831, Jordan
[3]School of Science, Technology and Engineering, University of Granada, 18010 Granada, Spain
[4]Information Systems and Networks Department, Faculty of Information Technology, The World Islamic Sciences and Education University, Amman 11947, Jordan
[5]Research Centre for Information and Communications Technologies of the University of Granada (CITIC-UGR), University of Granada, 1, 18010 Granada, Spain

Corresponding author: Hossam Faris (hossam.faris@ju.edu.jo)

ABSTRACT Online media has an increasing presence on the restaurants' activities through social media websites, coinciding with an increase in customers' reviews of these restaurants. These reviews become the main source of information for both customers and decision-makers in this field. Any customer who is seeking such places will check their reviews first, which usually affect their final choice. In addition, customers' experiences can be enhanced by utilizing other customers' suggestions. Consequently, customers' reviews can influence the success of restaurant business since it is considered the final judgment of the overall quality of any restaurant. Thus, decision-makers need to analyze their customers' underlying sentiments in order to meet their expectations and improve the restaurants' services, in terms of food quality, ambiance, price range, and customer service. The number of reviews available for various products and services has dramatically increased these days and so has the need for automated methods to collect and analyze these reviews. Sentiment Analysis (SA) is a field of machine learning that helps analyze and predict the sentiments underlying these reviews. Usually, SA for customers' reviews face imbalanced datasets challenge, as the majority of these sentiments fall into supporters or resistors of the product or service. This work proposes a hybrid approach by combining the Support Vector Machine (SVM) algorithm with Particle Swarm Optimization (PSO) and different oversampling techniques to handle the imbalanced data problem. SVM is applied as a machine learning classification technique to predict the sentiments of reviews by optimizing the dataset, which contains different reviews of several restaurants in Jordan. Data were collected from Jeeran, a well-known social network for Arabic reviews. A PSO technique is used to optimize the weights of the features, as well as four different oversampling techniques, namely, the Synthetic Minority Oversampling Technique (SMOTE), SVM-SMOTE, Adaptive Synthetic Sampling (ADASYN) and borderline-SMOTE were examined to produce an optimized dataset and solve the imbalanced problem of the dataset. This study shows that the proposed PSO-SVM approach produces the best results compared to different classification techniques in terms of accuracy, F-measure, G-mean and Area Under the Curve (AUC), for different versions of the datasets.

INDEX TERMS Sentiment analysis, SVM, PSO, SMOTE, oversampling, feature extraction, features weighting.

I. INTRODUCTION

The popularity of social media websites has witnessed tremendous growth in the last few years [1]. Social media

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

sites have grown not only in terms of volume but also in their importance to different aspects of life, including business, politics, and education [2]. Nowadays, all businesses are offering their products and services online. These sites allow consumers to share their experiences and recommendations about these businesses' products, places, and services on

different platforms such as TripAdvisor, Yelp, Facebook and Jeeran [3].

Online reviews represent the electronic version of word of mouth (WOM), which is an important aspect of in traditional marketing. While WOM is restricted to family, friends or close people, online reviews have a worldwide reach [4]. Many websites allow users to rate and review different products and services. These reviews become the main source of information for potential customers who are seeking such products [5]. A survey conducted by BrightLocal (2020) found that 79% of customers trust online reviews as much as personal recommendations [6].

These days, whenever a customer wants to buy a new product online, he or she will consider what other people think about it, how they rate it, and their feedback and comments about the product before making a purchase [7]. According to a BrightLocal survey in 2020, 87% of consumers had checked online reviews of local businesses [6]. These reviews may affect the customer's final choice since people trust customers' reviews more than advertisements produced by a company. Furthermore, customers' experiences can be enhanced by utilizing other customers' suggestions [3].

Due to the widespread availability of social websites and applications, the number of reviews available for various products has dramatically increased [8], and so has the need for automated methods to collect and analyze these reviews [9]. These methods are essential to speed up and improve the quality of decision making process [10].

Sentiment Analysis (SA) can be used to deduct users' feelings about various topics by processing their implicit attitudes and analyzing the underlying sentiments hidden in their comments [11]. Sharif *et al.* [12] defined SA as "analyzing people's sentiments, opinions, appraisals, attitudes, evaluations and emotions towards such entities as organizations, products, services, individuals, topics, issues, events and their attributes, as presented online via text, video and other means of communication." Sentiment analysis is also referred to as opinion mining based on natural language processing, text analysis, and computational techniques [13]. It can be applied at the document level, sentence level, or aspect level [14]. It aims to classify customers' attitudes towards a product or service as expressed in the comments, reviews, and posts as positive, negative, or neutral comments [15]. Two main SA approaches can be followed, namely, machine learning approaches and the lexicon-based approach [16]. Different machine learning algorithms are used to evaluate results in the sentiment field; the most common ones are Naïve Bayes (NB), SVM, Logistic Regression (LR), Random Forest (RF), and K-Nearest Neighbors (k-NN) [13].

Sentiment analysis is essential for every business as it can be used to improve the decisions of customers, business owners, and service providers [17]. SA is used by business owners to enhance their businesses' image and increase their success [3] since it helps decision-makers improve the quality of their products and services based on their customers' reviews; thus, the business can provide more praiseworthy

services. This leads to higher customer satisfaction and more sales and revenues for the business [12]. On the other hand, customers can utilize these reviews in making thoughtful decisions based on previous customers' experiences [17].

Recently, it has been noticed that almost all restaurants have presence in the online social world. Restaurants are becoming increasingly present on different social websites, and so are customers' reviews of these restaurants [18]. Online restaurant reviews are considered a rich source of information that helps attract new customers. Checking reviews by locals and tourists before visiting restaurants has become a trend [17]. This is supported by 2020 BrightLocal survey revealing that 93% of consumers check a restaurant's reviews before visiting it [6]. Consequently, customers' reviews can influence the success of restaurant business [19]. It was found that the more positive comments a restaurant receives the more customers visit its web pages and physical locations, which leads to more popularity and success [4]. In contrast, negative comments lead to the loss of trustworthiness of the restaurant and reduced revenue [17]. According to 2020 BrightLocal survey, 94% of consumers are more likely to buy from a business if it has received positive reviews, while 92% are less likely to use it if it has been given bad reviews [6]. People tend to post reviews when they either a strong positive or strong negative experience (generally, the number of positive reviews exceeds the number of negative ones) [4].

Customers' reviews and opinions are considered the final judgment of the overall quality of any restaurant. Thus, owners need to analyze their customers' underlying sentiments so they can meet their expectations and offer customized services in terms of food quality, ambiance, price, and customer service [18].

Many studies have followed a Machine Learning (ML) approaches for restaurants sentiment analysis. A study done by Zahoor *et al.* [3] used NB Classifier, logistic regression, SVM, and RF methods to analyze customers' sentiments about restaurants in Karachi. The study annotated 4000 reviews from a well-known Pakistani Facebook community called SWOT'S. Random forest gained the highest performance, with an accuracy of 95%. Another study conducted by Sharif *et al.* [19] classified customers reviews for 1000 restaurants (written in Bengali) into positive and negative classes using three machine learning algorithms, namely, Decision Tree (DT), RF, and multinomial NB. The results showed that the multinomial NB method achieved the best results, with 80.48% accuracy.

Furthermore, sentiment analysis can be used to build a recommender system in different fields including the restaurant industry. Asani *et al.* [11] for example collected people's sentiments from the TripAdvisor website and built a customized restaurant recommender system based on people's opinions and food preferences. The recommender system suggests restaurants according to users' preferences, thus helping them to choose the best option and make an informed decision. Choosing the best restaurant among many unknown

options is an important decision, especially for tourists and travelers.

Few studies have followed an evolutionary approach in the restaurant field. Govindarajan [20], for example, applied a hybrid classification method based on restaurant reviews found on Yelp. The study used NB, SVM, and Genetic Algorithm (GA) and then compared their performances with the proposed hybrid model built by coupling all three classification methods. Another study conducted by Somantri *et al.* [21] proposed a hybrid model for restaurant culinary food reviews in Indonesia. The study confirmed the efficiency of PSO, as it used a hybrid model baed on Particle Swarm Optimization (PSO) and Information Gain (PSO-IG) with four different classification algorithms, namely SVM, NB, DT, and K-NN. The best results were achieved using the proposed PSO-IG method with the NB classifier. The main limitation of this work was that it ignored the imbalanced nature of the dataset, as positive reviews were significantly more common than negative reviews. All aforementioned studies were applied to English sentiment; to the best of our knowledge, our work is the first to use such recent evolutionary approach to explore Arabic sentiments.

This work uses the Jeeran website to collect people's comments about different restaurants in Jordan. The Jeeran website is a social platform on which customers can post their reviews about more than 300,000 different places, including shopping centers, cafes, restaurants, or doctors' offices. People use Jeeran to find the best places and services in their cities and avoid bad experiences. Customers post thousands of comments on different social platforms every day, mainly in the Arabic language.

This study is conducted on Arabic sentiment since it is the fifth-most widely spoken language in the world and the first language of more than 422 million people [22]. Moreover, about 185 million web users are Arabic speakers [23]. The Arabic language is a more challenging language to study than to English for many reasons. Firstly, Arabic has a dialectal variety; people often post their comments in dialectical Arabic rather than Modern Standard Arabic, thus requiring more complex prepossessing [24]. Another reason is the morphology of the Arabic language, meaning the same word may have a different meaning, even if it has the same root. Also suffixes, affixes, and prefixes added to the same word may carry essential information [23]. Moreover, the richness of synonyms in the Arabic language plays a key role in its complexity. Furthermore, the same word may have different meanings according to the context, and a word can fit into more than one lexical category [22].

In addition, the problem of imbalanced datasets is very common since the majority of the sentiments of customers fall into either the supporter or resistor category. Parameter tuning of the oversampling technique is another challenge. Moreover, a large number of features is generated by tokenizing the sentiments of the customers. Thus, leading methodologies can be applied to feature analysis to achieve the best outcome from the classification process. Optimization techniques are well-suited for feature analysis and parameter tuning, making the classification process more reliable. Another challenge relates to the choice of tokenization method used to form the dataset for the classification process, and it can be recognized through experimental practice.

This research proposes an evolutionary approach to analyzing people's sentiments regarding restaurants' reviews in the Arabic language. Furthermore, this work followed an evolutionary hybrid approach by combining the PSO evolutionary algorithm with different oversampling techniques and the SVM algorithm to automatically detect the sentiment in the customers' comments. Four different oversampling techniques are applied to handle the problem of imbalance in the dataset. Additionally, the applied evolutionary algorithm helps reduce the effort and time needed to tune the parameters and optimize the classification by finding the best feature weights and best *k* value for the oversampling technique, thereby resulting in better performance measures.

This work collects the reviews for almost 3000 restaurants from the Jeeran website. After the data preparation process, four different versions of the dataset are presented using different tokenization methods. The initial individual of the study is created of random weights and a random k value for the oversampling parameter. The weighted oversampled data is then classified using the SVM classification technique and the results are evaluated using G-mean. A Particle Swarm Optimizer (PSO) evolutionary algorithm is then used to optimize the values of the individual and achieve a better G-mean. Finally, the proposed approach is compared with different standard and powerful classification models, including SVM, XGBoost, DT, RF, NB, k-NN, and LR based on Accuracy, $F1_P$, $F1_N$, G-mean, and AUC evaluation measures.

The main contributions of the study can be summarized as follows:

- Collecting the dataset from the Jeeran website with approximately 3000 restaurant reviews. The dataset is cleaned, labeled, formatted, and stemmed.
- Oversampling the dataset using four different oversampling techniques to solve the imbalanced problem.
- Applying the PSO optimization technique to find the best weights for the dataset features and the best *k* value for each oversampling technique, then applying the SVM classification technique to the oversampled and weighted dataset to find the sentiments of the restaurant reviews.

The remainder of the paper is divided as follows: Section II presents a review of the literature on restaurants' sentiment analysis. Section III introduces the backgrounds of different methods and concepts that have been used. The proposed approach is described in Section IV. Section V discusses the experiments and provides the results achieved by the proposed approach and other models. Finally, the conclusion and future directions are are offered in Section VI.

II. RELATED WORK

Sentiment analysis (SA) is one of the most studied research areas combining natural language processing, data mining,

and web mining. Owing to its importance to business and community, SA research has spread into management and social sciences as discussed by Liu [25].

As explained by Tubishat *et al.* [26], SA, which is also referred to as opinion mining, is a text-classification field in which people's opinions, evaluations, attitudes, moods, and emotions regarding a service or product are analyzed to detect orientations. It is conducted computationally by using natural language processing, linguistics, or text analysis to detect the feelings expressed within informal text posted online. Recently, due to the popularity of social networks and online review websites, people tend to check a restaurant's reviews before visiting it. As a result, customer's impressions have become a vital factor influencing the success of restaurants; the interest among decision- makers toward customers' experience about services provided has also increased as stated by Sharif *et al.* [19]. SA has been applied to online reviews about restaurants in the literature. For instance, Gan *et al.* [5] studied the attributes representing consumers' reviews of restaurants. This study found that the attributes derived from previous studies such as food, service, ambiance, and price were not enough to affect restaurants' ratings and that context should be added as a significant attribute. Meanwhile, Aye and Aung [1] proposed a Myanmar language resource for lexicon-based SA as a solution to language-specific problems since most studies have considered the English language for SA. Restaurant review data were used, but informal expressions were not addressed.

Since online booking websites gained substantial interest recently, and since people now check hundreds of reviews before making any booking decisions, Agüero-Torales *et al.* [27] proposed a cloud-based software tool to analyze data from the TripAdvisor website by conducting SA on them in the province of Granada.

The SA task was accomplished by using various datasets, such as the Yelp dataset, to examine the approaches proposed by other researchers as explained by Hegde *et al.* [28] and made public for research and academic studies. The Zomato Restaurant Dataset is derived from the online multinational restaurant aggregator in which reviews are posted alongside information, menus, and delivery options. Also, Taneja *et al.* [29] discussed that Zomato is a very rich database that includes information on more than 20,000 restaurants. Zomato API enables users to access the most up-to-date content and generate information about nearby restaurants. Furthermore, SemEval Datasets are high-quality annotated datasets generated through a series of international workshops; different versions of these datasets (e.g., SemEval-2015, SemEval-2016) have been used in literature as stated by Khan *et al.* [30].

SA is conducted through three main approaches, namely, the lexicon, machine learning (ML), and hybrid approaches. However, the application of lexicon-based SA has several drawbacks. Firstly, it requires a massive number of linguistic resources. Secondly, a predefined list of polarity annotation is required, and it differs based on the language used. Thirdly,

most written reviews use informal words that are not likely included in the lexicon. As a result, ML is a preferable alternative since it includes highly illustrative and flexible models, as stated by Alaei *et al.* [31].

Various ML algorithms have been used to conduct SA in the restaurants domain. NB, LR, and DT. ML algorithms were applied by Hassan *et al.* [32] to conduct SA on three different datasets, namely, the Yelp dataset, IMDB dataset, and Arabic qaym.com restaurant reviews dataset. Performance was measured in terms of accuracy and recall. NB and LR recorded the best results. Similarly, NB, SVM, multilayer perceptron, DT, k-NN, and fuzzy logic were applied by Kumar and Jaiswal [33] on data extracted from Twitter and Tumblr, which are widely used micro-blogging social networks. A comparative analysis of performance is presented in terms of precision, recall, and accuracy. Besides, a deep learning model called DOC-ABSADeepL was proposed by Zuheros *et al.* [34] and applied on the TripAdvisor dataset for restaurants to categorize the aspects included in an expert review while also extracting opinions and criteria. The tripR-2020 dataset was built, manually annotated, and released before being used in the same study.

Several implementations have considered the Arabic language for conducting SA using ML. For instance, Al Omari *et al.* [35], logistic regression (LR) was applied on data extracted from reviews (including restaurant reviews) posted on Google and Zomato about public services in Lebanon. Several ML algorithms, namely, KNN, NB, SVM, LR, and RF, were applied for SA by Alharbi and Qamar [36] to assess customers' reviews about restaurants and cafes in the Qassim region of Saudi Arabia. Performance was measured based on accuracy, recall, and F-measure, with the best performances for these measures produced by SVM, LR, and RF, respectively.

Class imbalance problem have been considered in some sentiment analysis studies, such as Qiu *et al.* [37] in which a heuristic re-sampling algorithm was applied as solution to imbalance data encountered while training the proposed model. While Pongthanoo and Songpan [38] proposed a technique which combines Information Gain (IG) with SMOTE to improve performance accuracy of ML classifiers when applied on imbalanced dataset. A GAN-SMOTE architecture was proposed by Scott and Plested [39] in which generative adversarial networks (GANs) was merged with SMOTE in order to up-sample their dataset and generate convincing synthetic examples of it for improving performance accuracy of their experiments. SMOTE technique was also applied by Nguyen *et al.* [40] on restaurant data crawled from Foody, an online community people use to search and comment on food in Vietnam, to overcome the class imbalance challenge for supervised learning method to be applied afterward.

Evolutionary algorithms (EAs) were applied during the pre-processing step to select the most effective attributes for building the model. For example, a hybrid SA framework applying the GA feature reduction approach was proposed by Iqbal *et al.* [41] in which both principal

component analysis (PCA) and latent semantic analysis (LSA) were applied for comparison purposes with the proposed GA-based approach, which outperformed them.

Support vector machine (SVM) techniques have been applied as a classifier after the feature-selection step using an EA. The study by Kumar and Khorwal [42] applied a support vector machine (SVM) for classification after the feature-selection step using the firefly algorithm to find the optimum subset of features. Experiments were conducted on four datasets (two English and two Hindi), and a genetic algorithm (GA) was applied for comparison purposes. The whale optimization algorithm (WOA), which is one of the recent metaheuristic algorithms introduced by Tubishat *et al.* [43] to solve the problem of falling in local optima, was improved by including elite opposition-based learning (EOBL) in the initialization step and including evolutionary operators, such as mutation, crossover, and selection operators, at the end of each iteration. Along with the filter feature selection technique, information gain (IG) was considered using SVM. The proposed improvements were validated using four Arabic datasets for SA while six ML and two DL algorithms were applied.

Generally, we noticed that the majority of the previous works that proposed machine learning techniques for sentiment analysis in the Arabic language context focused on applying classical supervised classification techniques like SVM, NB and Decision trees. However, there are less works tried to solve the class imbalance challenge in the data distribution. Therefore, in this work, we are going to follow a different line of research by integrating Evolutionary Algorithms (EA) with different variants of the SMOTE oversampling techniques and SVM. The goal of this integration is to overcome the class imbalance challenge and at the same time improving the classification power for the targeted SA problem in the Arabic language context.

III. PRELIMINARIES
A. PARTICLE SWARM OPTIMIZATION
Particle swarm optimization (PSO) is a swarm intelligence algorithm developed for solving nonlinear problems within various sciences and engineering domains; it was derived from the flocking of birds or schooling of fish, as stated by [44].

PSO is a search algorithm that uses swarm intelligence to find solutions as explained in [45] and [46]. It generates a random search result by analyzing a set of potential solutions (called a *swarm*); every single potential solution is referred to as a *particle*. Particles normally rely on two kinds of learning while moving: cognitive learning and social learning. The former refers to the process of learning from other particles (the result is stored as *gbest*), while the latter refers to the process of storing the best solution that was visited by the particle (stored *pbest*). Velocity is used to determine the magnitude and direction of a particle. It refers to changing position rate with respect to time—that is, an iteration in the case of PSO.

Taking velocity as v and position as x, both are the same as counter increases by unity for iterations. Equation 1 describes the velocity update.

$$v_{id}^{t+1} = v_{id}^t + c_1 r_1 (p_{id}^t - x_{id}^t) + c_2 r_2 (p_{gd}^t - x_{id}^t) \qquad (1)$$

where in a D-dimensional search space, D-dimensional vector is denoted as $x_i^t = (x_{i1}^t, x_{i2}^t, \ldots, x_{iD}^t)^T$ represents *i*th particle of the swarm at *t* time step. Velocity is represented as $v_i^t = (v_{i1}^t, v_{i2}^t, \ldots, v_{iD}^t)^T$. The best-visited position previously of the *i*th particle of swarm at *t* time step is denoted as $p_i^t = (p_{i1}^t, p_{i2}^t, \ldots, p_{iD}^t)^T$ and index of best particle in the swarm is referred to as *g*, while c_1, c_2 are constants representing cognitive and social scaling parameters, r_1, r_2 are random numbers in the range [0, 1].

The position is updated as in equation 2, where d is the dimension and i is the particle index. Algorithm 1 illustrates the pseudo-code of PSO.

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \qquad (2)$$

Algorithm 1 Pseudo-Code of Particle Swarm Optimizer

1 Create and Initialize a D-dimensional swarm, S and corresponding velocity vectors
2 *for* (*t* = 1 to the maximum bound on the number of iterations) *do*
3 *for*(*i* = 1 to *S*) *do*
4 *for*(*d* = 1 to *D*) *do*
5 Apply the velocity update equation
6 Apply position update equation
7 *end*
8 Compute fitness of updated position
9 If needed, Update historical information for pbest and gbest
10 *end*
11 Terminate if gbest meets problem requirements
12 *end*

B. SUPPORT VECTOR MACHINE
The support vector machine (SVM) algorithm is a supervised classifier that is applied widely to solve classification and regression problems. It was designed as an improvement to the support vector classifier, which has been introduced as an enhancement to the maximal margin classifier, which is restricted to dealing with simple linearly separable data [47].

In high-dimensional vector spaces where the feature space is well-divided, SVM generates linearly separated hyperplanes. These planes partition the data points belonging to the two classes into distinct regions. The optimal hyperplane is always the one that maximizes the distance between the nearest training data points and the feature space [48].

Significant misclassification can appear because of linear classification since data points belonging to distinct classes are rarely clearly distinguished. SVM can handle such commonly faced cases, as it maps the feature space into a higher-dimensional space where non-linear data points are

transformed into linearly separable points. Therefore, are transformed into linearly separable points. Therefore, data points that belong to different classes have clearer separation boundaries [49].

The kernel function is used to generate the high-dimensional space by enlarging the original space non-linearly. There are several forms of kernel functions, namely linear kernels, polynomial kernels, and radial basis function kernels. Equations 3, 4, and 5 describe them respectively, where k () denotes the kernel function and the product of the two observation vectors x_i and x_i' represents its outcome. The two vectors' product is referred to as $\varphi(x_i).\varphi(x_i')$, where φ is the transformed feature space [50].

$$k\left(x_i, x_i'\right) = \sum_{j=1}^{p} x_{ij} x_{ij}' \tag{3}$$

$$k\left(x_i, x_i'\right) = (1 + \sum_{j=1}^{p} x_{ij} x_{ij}')^d \tag{4}$$

$$k\left(x_i, x_i'\right) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{ij}')^2) \tag{5}$$

C. OVERSAMPLING TECHNIQUES

In classification problems, having the target class label unequally distributed causes a situation that is commonly encountered. These data can be referred to as an imbalanced dataset, which affects the training process of the data mining model, as it will be conducted mainly on the majority class, causing bias in class predictions, as the minority class holding few instances may be considered as noise or outliers. As a result, imbalanced datasets pose serious challenges by affecting classifiers' performance, as explained by [51].

Hence, solving data imbalance issues is vital and should be conducted as a preliminary step before classification as was done in [52]. Various balancing techniques are applied in this regard. They can be categorized into oversampling techniques—such as SMOTE and adaptive synthetic sampling (ADASYN)—and undersampling techniques, such as edited nearest neighbors, random under sampling, and TomekLinks. The former refers to the artificial creation of minority class points in the dataset, while the latter removes the majority class labels from the dataset.

1) SMOTE

The synthetic minority oversampling technique (SMOTE) [53] is an oversampling technique that is widely applied in data mining to solve imbalance datasets as explained by [54]. It focuses on the "feature space" instead of the "data space," as it does not replicate minority classes but instead introduces synthetic instances by randomly choosing a minority class and then interpolating their K-nearest neighbors. KNN generates instances within the dataset by considering other instances near them since it is naturally applied to find the closest neighbors of a specific point [55].

The process starts by setting up N, which is the total amount of oversampling (presented as an integer value). This is done either by approximating a 1:1 class distribution or using the wrapper process to discover the class distribution. Then, an iterative set of steps are carried out, starting with randomly selecting a minority class instance training set. After that, the K nearest neighbors are obtained, with the value of k set to 5 by default. Finally, new instances are computed by selecting N K-NN instances.

The final step is obtaining the difference between the feature vector of the sample being processed and every single selected neighbor. Afterward, this difference is multiplied by a random number between 0 and 1; the result is then added to the previous feature vector. This results in the selection of a random point within the line segment of features. Eventually, one of the nominated attributes is selected. The entire process is summarized in Algorithm 2.

Algorithm 2 Pseudo-Code of SMOTE Algorithm

---

1 *function* **SMOTE(T;N; k)**
  Input: T; N; k   #minority class examples, Amount of oversampling, #nearest neighbors
  *Output: (N/100) * T*   synthetic minority class samples
  Variables:
  Sample: array for original minority class samples
  newindex: keeps a count of number of synthetic samples generated, initialized to 0
  Synthetic: array for synthetic samples
2 **if** $N < 100$ **then**
3 | Randomize the T minority class samples
4 | $T = (N/100) * T$
5 | $N = 100$
6 **end**
7 $N = (int)N/100$ The amount of SMOTE is assumed to be in integral multiples of 100
8 **for** $i = 1$ *to T* **do**
9 | Compute k nearest neighbors for i, and save the indices in the *nnarray*
10 | POPULATE($N; i; nnarray$)
11 **end**
12 *end function*

---

2) SVM-SMOTE

SVM-SMOTE is a variant of the SMOTE algorithm that deploys an SVM classifier to capture a sample to be used for new synthetic samples generation. The process is conducted by applying SVM to the original training dataset after approximating the borderline area using support vectors. The algorithm emphasizes data separation, as it synthesizes data far from class overlap. Data is synthesized randomly along lines joining every minority class support vector by referring to its nearest neighbors [56].

3) ADAPTIVE SYNTHETIC SAMPLING ADASYN

Adaptive synthetic sampling (ADASYN) is another variation of SMOTE that differs from focusing on neighbors or

borderlines. Instead, it focuses on data density and creates synthetic data accordingly [52].

Generating synthetic data is inversely related to minority class density. This means more synthetic data is generated in places within feature space where minority class density is low, while few (or no) such data are generated where minority class density is high [57]. In other words, where the minority class is less dense within the feature space, the synthetic data are created at a higher frequency; otherwise, no synthetic data are created.

4) BORDERLINE-SMOTE

As the name implies, borderline-SMOTE is a version of SMOTE that differs in functionality. Instead of creating synthetic data randomly related to near data, borderline-SMOTE tends to specify each class's borderline. Instances on the borderline and close to it are more likely than others to be misclassified than those far from borderline and, therefore, are more important for the classification task [58].

In borderline-SMOTE, all minority class instances are divided into three groups: noise that is rare, incorrect, and located in areas of majority class instances; danger instances, which are located on class boundaries and overlap with the majority class; and safe instances, which represent the minority class.

D. FEATURE EXTRACTION

Feature extraction (FE) is defined as the process by which a set of initial raw data is split dimensionality into groups that can be processed easily, as described by [59].

According to [60], FE is one of the most commonly applied techniques for reducing the dimensionality of data, as high-dimensional data is mapped into low-dimensional but potential features. These techniques extract only informative features, the use of which would cause a significant improvement to ML models' performance.

1) N-GRAM

The formation of text features for a supervised ML classifier can be done by N-gram, as stated in [61], where given text can be represented by several sequences of tokens n. If the value of n is 1, it is called unigram; if the value of n is 2, it is called bigram; if the value of n is 3, it is called trigram. For example, if the sentence "Jordan is better country" is considered and $n = 2$, then it will produce "Jordan is," "is better," and "better country."

2) BAG-OF-WORDS

A very popular technique for FE, as stated in [62], involves columns that represent words and columns that represent the value of a weight measure such as term frequency and term frequency-inverse document frequency. Meanwhile, in [63], it was discussed that bag-of-words (BoW) embraces features representing documents as vectors of words from a vocabulary. In other words, the BoW model includes a representation of every single document from the corpus in the form of

TABLE 1. Details of the datasets.

Datasets	Positive	Negative
Our-data	2150	640
Public-data	3465	451

vectors, along with the frequency of each word from a certain vocabulary. It has been used in SA applications as a robust technique despite its simplicity.

IV. METHODOLOGY

The methodology of this paper is described and presented in this section. Three phases have been provided in detail, including, data description, collection and preparation, and the proposed approach.

A. DATA DESCRIPTION AND COLLECTION

The datasets utilized in this work describe customers' reviews of various restaurants in Jordan. Data were collected from Jeeran, a well-known social network for Arabic reviews. Since 2010, the Jeeran website has provided a platform for comparing and reviewing the platform for comparing and reviewing the best places and services in the Arab world, including cafes, hotels, restaurants, and public services. Reviews of these types can provide useful feedback to those who make decisions about the quality of service and food, prices, and other aspects related to the ambiance of these places.

Approximately 3000 restaurant reviews have been compiled from the Jeeran website. The collection process is performed using a developed C# script that can easily and simultaneously be gathered from different pages of several restaurants' websites. Further, these collected reviews are stored as text files for further analysis. Furthermore, Table 1 shows the details of the datasets.

B. DATA PREPARATION AND LABELING

The collected dataset is cleaned, labeled, formatted, and stemmed [64], [65]. The cleaning process is performed on the dataset by removing symbols and special characters. [66], [67].

To label the reviews, we implemented a crowdsourcing website, where all reviews are uploaded and arranged for labeling. We invited more than 290 individuals to annotate the reviews, where 10 reviews were assigned for each person and labeled positive or negative. The reviewers were asked to read each review carefully and label it according to their understanding of the sentiment of that review. Two options were available under each review—negative and positive—for the reviewers to select. Consequently, the class of the review was assigned based on the majority of reviewers' selections.

Thereafter, each review was stored on its own file, and all files of the reviews were collected and stored according to the class type. A CSV file was created from all reviews, including their context in one column and their class label in the second.

TABLE 2. Different versions of tokenized datasets.

Datasets	Tokenization	# of Features
Data 1	1-Gram	3439
Data 2	2-Gram	8985
Data 3	3-Gram	14233
Data 4	Bag-of-Words	2916

The dataset comprises 2150 positive reviews and 640 negative reviews.

After dataset labeling, the formatting process is started. First, the stop words such as جدا ، ثم ، الذي ، إذا ، قد أنا ، هذا ، (translated to I'm, so, that, then, very, this and may respectively) are removed. Stop words have to be deleted since they do not affect the meaning of the text. After that, non-Arabic letters and emoticons are eliminated through a normalization process. Several useless features are eliminated using text normalization and stop word removal, which reduces the overall number of extracted features and enhances the feature selection process.

Finally, the stemming process is applied by the Lovins stemmer technique to remove duplicate words and supplementary letters. Then, N-gram and BOW feature extraction methods are applied for text tokenization, generating four different versions of the data (Table 2) As observed in the table, a different number of features is generated by each combination of feature extraction and stemming. The numbers of features are 3439, 8985, 14,233, and 2916 for the 1-gram, 2-gram, 3-gram, and BoW techniques, respectively.

C. PROPOSED APPROACH

Figure 1 represents the approach proposed in this paper and shows the steps applied to conduct the experiments and achieve the obtained results. After the preparation of the data, the dataset is split into training and testing parts. Then, the evolutionary classification of the training dataset is applied to the training part. Based on this process, the classification of the optimized weights and the k value are applied to the testing data.

Specifically, the training process, which can be observed in the left part of Figure 1, considers the creation of an initial individual consisting of random weights and a random k value for the oversampling parameter. The random weights of the initial individual are applied to the training data to generate weighted features' values for each instance, thus generating the weighted training data. Figure 2 shows the weighting process by which the weight part of the individual is multiplied by each instance to generate the weighted dataset. On the other hand, the weighted training data are used in the oversampling technique, with the k parameter value used as the random value obtained from the initial individual, to generate oversampled and weighted training data. The data are then classified using the SVM classification technique and evaluated by the fitness function generated in terms of G-mean. PSO is then used to optimize the values of the individual to generate a better G-mean value following

the same process applied to the initial individual. Multiple iterations are performed to optimize the classification to find a better weight and k value for the process. For the final iteration, the best individual is kept for the testing process.

The testing process, which is observed in the right part of Figure 1, uses the optimized individual to generate a weighted oversampled testing data. That is, the testing process generates the weighted testing data based on the weights of the individual and performs the oversampling process based on the k value of the individual. The classification of the data by SVM is then evaluated using the evaluation measures considered in this paper to evaluate the performance of the proposed approach.

V. EXPERIMENTS AND RESULTS

The experiments and their results are described and analyzed in detail in this section. Several stages are involved in this phase, including the experimental set-up, the evaluation measures, the results of PSO-SVM with different oversampling techniques, a comparison with some standard classification algorithms followed by a comparison against other recent studies approaches, and finally a feature importance analysis is discussed.

A. EXPERIMENTAL SETUP

In this study, the experiments were conducted on a PC running Windows 10 with a 2.40-GHz Intel Core i7 and 16 GB of RAM. Further, the scikit-learn library and EvoloPy framework were used to run tests in the Python environment. Our proposed approach settings were 100 iterations, a population size of 100, and 30 runs.

B. EVALUATION MEASURES

The performance of our model was evaluated by considering accuracy, F1-score (positive), F1-score (negative), g-mean, and AUC measures. Accuracy provides the classification quality of the model, calculated as the ratio between the true negatives (TNs) and true positives (TPs), as well as between TNs and false positives (FPs) based on Equation 6.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The F-measure is the consonant mean of precision and recall, and it can be calculated using the following equation:

$$F1_P = 2 * \frac{precision_P \times recall_P}{precision_P + recall_P} \tag{7}$$

$$F1_N = 2 * \frac{precision_N \times recall_N}{precision_N + recall_N} \tag{8}$$

G-mean is a measure of the balance of the classification performance between two classes. G-means are calculated mathematically by multiplying both recalls; recall-negative (REC_N) and recall-positive (REC_P) by the square root.

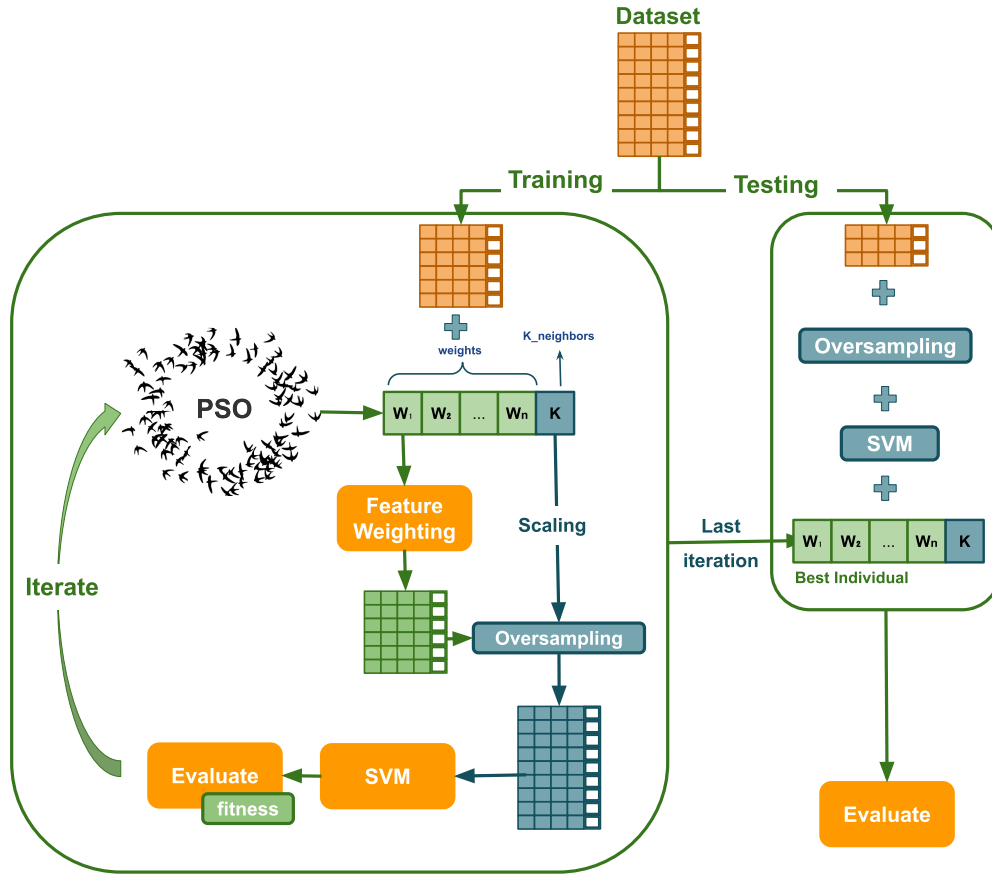$$G - mean = \sqrt{REC_N \times REC_P} \tag{9}$$

FIGURE 1. An illustration of the proposed PSO-SVM approach based on oversampling techniques.
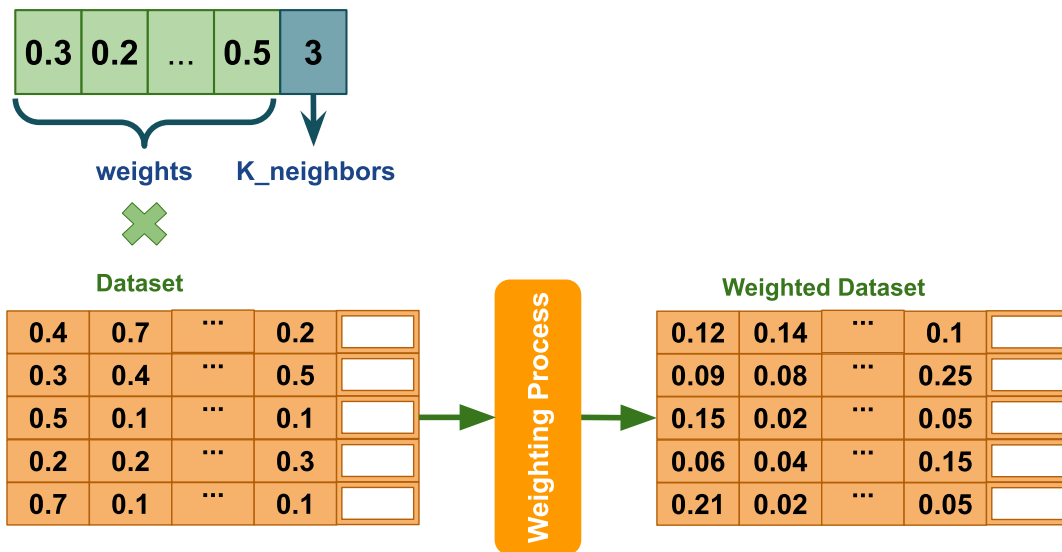

FIGURE 2. Illustration of the proposed weighting process utilized in the proposed PSO-SVM approach.

AUC is considered a well-established evaluation measure for classifiers. Random classifiers will have an AUC equal to 0.5, whereas a perfect classifier will have an AUC equal to 1.

$$AUC = \int_0^1 \frac{TP}{P} \, \mathrm{d}\frac{FP}{N} = \frac{1}{P \cdot N} \int_0^1 TP \, \mathrm{d}FP \qquad (10)$$

C. PSO-SVM WITH DIFFERENT OVERSAMPLING TECHNIQUES RESULTS

This phase explains the performance of the proposed PSO-SVM approach with different oversampling techniques and its comparison against all dataset versions. The considered oversampling techniques are SMOTE, SVMSMOTE,

TABLE 3. Accuracy, F-measure, and G-mean results for the proposed PSO-SVM for all datasets with four different oversampling techniques.

Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean
Data 1				
SVM-PSO+SMOTE	0.8794	0.9282	0.6196	0.7952
SVM-PSO+SVMSMOTE	0.8926	0.9368	0.6431	0.7974
SVM-PSO+ADASYN	0.8859	0.9343	0.6337	0.7943
SVM-PSO+BorderlineSMOTE	**0.8970**	**0.9396**	**0.6507**	**0.8006**
Data 2				
SVM-PSO+SMOTE	0.8810	0.9291	0.6283	**0.8044**
SVM-PSO+SVMSMOTE	0.8950	0.9384	0.6423	0.7898
SVM-PSO+ADASYN	**0.8988**	**0.9408**	0.6505	0.7947
SVM-PSO+BorderlineSMOTE	0.8978	0.9401	**0.6528**	0.8007
Data 3				
SVM-PSO+SMOTE	0.8836	0.9305	0.6395	0.8147
SVM-PSO+SVMSMOTE	**0.8999**	**0.9412**	**0.6623**	0.8072
SVM-PSO+ADASYN	0.8955	0.9385	0.6499	0.8029
SVM-PSO+BorderlineSMOTE	0.8964	0.9388	**0.6623**	**0.8170**
Data 4				
SVM-PSO+SMOTE	0.8738	0.9250	0.6005	0.7807
SVM-PSO+SVMSMOTE	**0.8856**	**0.9327**	**0.6190**	0.7805
SVM-PSO+ADASYN	0.8821	0.9306	0.6063	0.7732
SVM-PSO+BorderlineSMOTE	0.8850	0.9323	0.6186	**0.7844**

TABLE 4. Accuracy, F-measure, G-mean, and AUC results for Data 1 for the proposed PSO-SVM against other algorithms.

Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean	AUC
SVM-PSO+BSMOTE	**0.89**	**0.93**	**0.65**	**0.80**	**0.81**
SVM	0.85	0.91	0.56	0.79	0.79
XGBoost	0.76	0.85	0.45	0.73	0.74
DT	0.81	0.89	0.40	0.63	0.67
RF	0.83	0.90	0.31	0.51	0.60
NB	0.80	0.88	0.43	0.68	0.70
KNN	0.43	0.51	0.32	0.58	0.65
LR	0.85	0.91	0.56	0.76	0.77

TABLE 6. Accuracy, F-measure, G-mean, and AUC results for Data 3 for the proposed PSO-SVM against other algorithms.

Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean	AUC
SVM-PSO+BSMOTE	**0.89**	**0.93**	**0.66**	**0.81**	**0.82**
SVM	0.85	0.91	0.56	0.76	0.77
XGBoost	0.77	0.85	0.44	0.73	0.73
DT	0.82	0.89	0.43	0.66	0.69
RF	0.83	0.90	0.30	0.48	0.59
NB	0.87	0.92	0.56	0.74	0.75
KNN	0.45	0.53	0.32	0.58	0.65
LR	0.85	0.91	0.57	0.79	0.80

TABLE 5. Accuracy, F-measure, G-mean, and AUC results for Data 2 for the proposed PSO-SVM against other algorithms.

Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean	AUC
SVM-PSO+BSMOTE	**0.88**	**0.92**	**0.62**	**0.80**	**0.81**
SVM	0.85	0.91	0.56	0.77	0.78
XGBoost	0.77	0.85	0.44	0.72	0.73
DT	0.82	0.89	0.43	0.66	0.68
RF	0.83	0.90	0.30	0.49	0.59
NB	0.86	0.92	0.53	0.71	0.74
KNN	0.44	0.53	0.32	0.58	0.65
LR	0.85	0.91	0.57	0.78	0.79

TABLE 7. Accuracy, F-measure, G-mean, and AUC results for Data 4 for the proposed PSO-SVM against other algorithms.

Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean	AUC
SVM-PSO+BSMOTE	**0.88**	**0.93**	**0.61**	**0.78**	**0.79**
SVM	0.76	0.85	0.43	0.71	0.78
XGBoost	0.77	0.86	0.48	0.75	0.71
DT	0.80	0.88	0.38	0.62	0.65
RF	0.83	0.90	0.30	0.49	0.59
NB	0.79	0.87	0.40	0.65	0.67
KNN	0.25	0.22	0.27	0.35	0.55
LR	0.85	0.91	0.57	0.78	0.79

ADASYN, and borderline-SMOTE. Table 3 shows the results of the datasets in terms of accuracy, F-measure, and g-mean. All datasets are presented in ascending order, from Data 1 to Data 4.

As shown in the first part of the table, the highest results for Data 1 obtained by SVM-PSO+BorderlineSMOTE in terms of accuracy, $F1_P$, $F1_N$, and g-mean were 0.897, 0.939, 0.650, and 0.800, respectively.

As for Data 2, the SVM-PSO+ADASYN outperforms the other algorithms in terms of accuracy and $F1_P$. Meanwhile, for $F1_N$, the best results were obtained by

SVM-PSO+BorderlineSMOTE; for g-mean, the best results were obtained by SVM-PSO+SMOTE.

In the third part (Data 3), the best accuracy and $F1_N$ were achieved by SVM-PSO+SVMSMOTE. For $F1_N$, SVM-PSO+SVMSMOTE and SVM-PSO+Borderline SMOTE both acquired the highest result of 0.662. Meanwhile, the best g-mean was obtained by SVM-PSO+ BorderlineSMOTE.

Further, the SVM-PSO+SVMSMOTE performed better than the other algorithms for Data 4 in terms of accuracy, $F1_P$ and $F1_N$ (0.8856, 0.9327, and 0.6190, respectively).

TABLE 8. Accuracy, F-measure, G-mean, and AUC results for the proposed PSO-SVM against recent studies.

Token.	Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean	AUC
1-gram	SVM-PSO+BSMOTE	0.89	0.93	0.65	0.80	0.81
2-gram	SVM-PSO+BSMOTE	0.88	0.92	0.62	0.80	0.81
3-gram	SVM-PSO+BSMOTE	0.89	0.93	**0.66**	**0.81**	**0.82**
BoW	SVM-PSO+BSMOTE	0.88	0.93	0.61	0.78	0.79
TF-IDF	Bidirectional LSTM	0.86	0.93	0.13	0.26	0.53
TF-IDF	GBDT	**0.91**	**0.95**	0.57	0.65	0.71

TABLE 9. Accuracy, F-measure, G-mean, and AUC results for OCLAR dataset for the proposed PSO-SVM against other algorithms.

Algorithms	Accuracy	$F1_P$	$F1_N$	G-mean	AUC
SVM-PSO+BSMOTE	**0.80**	**0.88**	**0.39**	**0.68**	**0.71**
SVM	0.76	0.86	0.35	0.66	0.67
XGBoost	0.66	0.77	0.33	0.68	0.68
DT	0.73	0.83	0.28	0.60	0.61
RF	0.75	0.85	0.28	0.59	0.62
NB	0.50	0.63	0.23	0.56	0.56
KNN	0.59	0.71	0.30	0.66	0.67
LR	0.77	0.86	0.37	0.68	0.69

However, in terms of g-mean, the SVM-PSO+Borderline SMOTE achieved the highest result (0.7844).

It is worth mentioning that the PSO-SVM yielded the best g-mean results, which was considered the main measurement in this study since the datasets are imbalanced. It is essential to ensure that the classification approach makes quality predictions for the majority class and other classes.

D. STANDARD CLASSIFICATION MODELS COMPARISON

In this subsection, the experiments focused on comparing the proposed PSO-SVM with different standard classification models, including SVM, XGBoost, DT, RF, NB, k-NN, and LR. All these standard classification models were also combined with an oversampling technique. Table 4 shows that the SVM-PSO+BorderlineSMOTE outperformed all other algorithms on Data 1 in terms of all measures, while the standard SVM ranked second and the LR placed third.

As shown in Table 5, the highest results were acquired by SVM-PSO+SMOTE in all measures; however, the NB was the second-best algorithm in terms of accuracy and $F1_P$, with values of 0.86 and 0.92, respectively. As for $F1_N$, g-mean, and AUC, the LR obtained the second-best results of 0.57, 0.78 and 0.79, respectively. Meanwhile, the standard SVM yielded values of 0.85, 0.91, 0.56, 0.77, and 0.78 in terms of accuracy, $F1_P$, $F1_N$, g-mean, and AUC, respectively.

The results for Data 3 show similar outcomes, with the SVM-PSO+BorderlineSMOTE emerging as the superior algorithm, with results of 0.89, 0.93, 0.66, and 0.81 in terms of accuracy, $F1_P$, $F1_N$, g-mean and AUC, respectively. The NB, LR, and standard SVM placed second, third, and fourth, respectively.

For Data 4, the SVM-PSO+BorderlineSMOTE also outperformed the other algorithms, with values of 0.88, 0.93, 0.61, 0.78, and 0.79 for accuracy, $F1_P$, $F1_N$, g-mean, and AUC, respectively. However, in this data, the NB's performance decreased in terms of accuracy, $F1_P$, $F1_N$, g-mean, and AUC when compared with the other algorithms.

Moreover, the LR achieved the second-best results for all measures, while the RF ranked third in terms of accuracy and $F1_P$; XGBoost ranked third for $F1_N$ and g-mean; while standard SVM ranked third in terms of AUC.

In order to demonstrate the performance of the proposed approach, two different comparisons have been added. In the first comparison the proposed approach is compared against other recent studies in the literature [68]. The used techniques in these studies were TF-IDF-Bidirectional-LSTM and TF-IDF-GBDT. As can be seen in table 8, the results illustrates that the proposed approach (SVM-PSO+BSMOTE) outperform the TF-IDF-Bidirectional-LSTM technique in all performance measurements. Moreover, our approach (3-gram-SVM-PSO+BSMOTE) achieved the highest results in terms of $F1_N$, g-mean, and AUC, while TF-IDF-GBDT obtained the best result for accuracy and $F1_N$. Since this work aims to solve the imbalance problem, G-mean and AUC are considered more important than the other measures.

Additionally, in the second comparison, the proposed approach was compared with other algorithms based on the Opinion Corpus for Lebanese Arabic Reviews (OCLAR) dataset that was published in the UCI repository (Table 9). The results support our research findings and prove the superiority of our proposed approach, since it can be noticed that the SVM-PSO+BSMOTE achieved the best results in all measures.

In summary, the PSO-SVM with BorderlineSMOTE and SMOTE obtained the best results in all measures for all datasets, followed by NB, LR, and standard SVM (Tables 4, 5, 6, 7, and 9). While in comparison with recent studies the proposed PSO-SVM outperforms the other algorithms in terms of $F1_N$, g-mean, and AUC.

E. DISCUSSION

Four different versions of the dataset were presented in this study. Three of the versions were created using the N-gram method, while the fourth was created by using the bag-of-words method. The different data versions produced different numbers of features (Table 2).

In the first instance, the datasets were examined using the proposed PSO-SVM (Figure 3). According to the figure, Data 2 showed the highest results for accuracy and $F1_P$, while Data 3 presented the highest results in terms of $F1_N$ and g-mean.

Overall, the classification performance of all datasets improved as the number of features increased. Accordingly, the use of many features (N-grams) has a positive influence on classification performance. The second part of the analysis of the results revealed how poorly the bag-of-words method performed compared to other methods.

Additionally, Figure 4 illustrates the g-mean results for each algorithm across all datasets. The results clearly show that the classification performance of the standard SVM was improved by using PSO. By using feature weighting and optimizing the SVM parameters, the PSO-SVM achieved better results. Thus, the PSO improved the results of the standard method by 0.01, 0.03, 0.05, and 0.07 for Data 1,
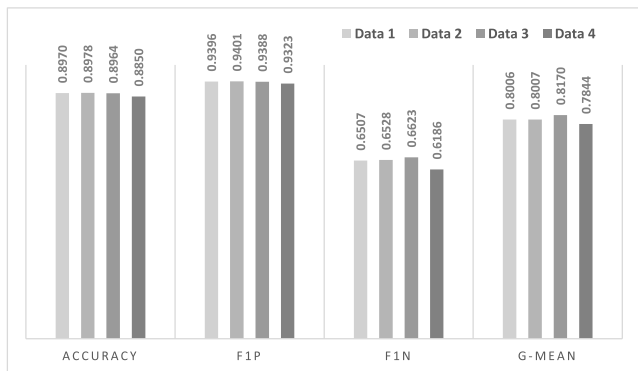
FIGURE 3. Comparison of results for all datasets for the proposed PSO-SVM.
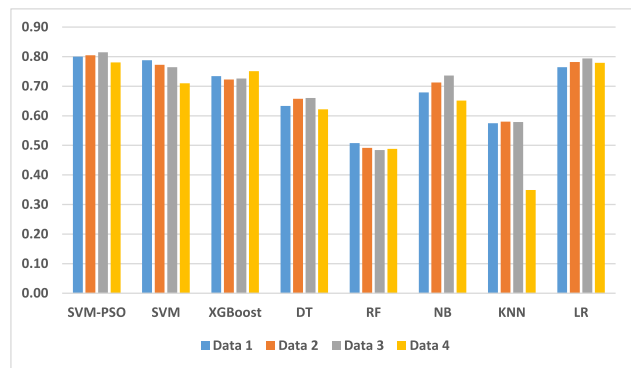

FIGURE 4. G-mean results of all datasets for all algorithms.

Data 2, Data 3, and Data 4. We also noticed that the proposed PSO-SVM outperformed the other algorithms in all datasets. On the other hand, the k-NN algorithm obtained the worst results due to the complexity of the data in terms of instances and dimensions.

VI. CONCLUSION

Sentiment analysis has witnessed increased interest in the academic field in the last few years. Many people post reviews of different services and products. The analysis of customers' attitudes and feedback is essential for all businesses, including restaurants. Thus, this research proposed a new hybrid evolutionary technique that aims to analyze people's sentiment towards various restaurants across Jordan. The data were collected from a popular social network, namely Jeeran. The proposed approach consisted of collecting more than 3000 restaurant reviews and labeling them using the crowdsourcing technique. Oversampling techniques were then applied to solve the problem of imbalanced data in the dataset. We produced four versions of the collected dataset using different tokenization methods, including 1-Gram, 2-Gram, 3-Gram, and bag-of-words. Further, we implemented a hybrid optimization technique comprising PSO and SVM to find the best weights while also finding the k values of four different oversampling techniques to predict the sentiments of reviews. The study demonstrates that the proposed PSO-SVM approach is effective and outperforms the other approaches in

all investigated measures (accuracy, F-measure, g-means, and AUC). In more detail, the PSO-SVM provided better results than the standard SVM, LR, RF, DT, k-NN, and XGBoost in all versions of the datasets. We plan to employ different metaheuristic algorithms on this data in future work. Moreover, other applications can be applied to predict the sentiments of reviews for other products, such as medical and engineering products.

REFERENCES

[1] Y. M. Aye and S. S. Aung, "Senti-lexicon and analysis for restaurant reviews of Myanmar text," *Int. J. Adv. Eng., Manage. Sci.*, vol. 4, no. 5, Jan. 2018, Art. no. 240004.

[2] P. P. Rokade and A. K. D, "Business intelligence analytics using sentiment analysis—A survey," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, p. 613, Feb. 2019.

[3] K. Zahoor, N. Z. Bawany, and S. Hamid, "Sentiment analysis and classification of restaurant reviews using machine learning," in *Proc. 21st Int. Arab Conf. Inf. Technol. (ACIT)*, Nov. 2020, pp. 1–6.

[4] M. Nakayama and Y. Wan, "The cultural impact on social commerce: A sentiment analysis on yelp ethnic restaurant reviews," *Inf. Manage.*, vol. 56, no. 2, pp. 271–279, Mar. 2019.

[5] Q. Gan, B. H. Ferns, Y. Yu, and L. Jin, "A text mining and multidimensional sentiment analysis of online restaurant reviews," *J. Quality Assurance Hospitality Tourism*, vol. 18, no. 4, pp. 465–492, Oct. 2017.

[6] R. Murphy. (Dec. 9 2020). *Local Consumer Review Survey 2020*. BrightLocal. Accessed: Nov. 5, 2021. [Online]. Available: https://www.brightlocal.com/research/local-consumer-review-survey/

[7] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, 2013.

[8] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6000–6010, 2012.

[9] L. Li, L. Yang, and Y. Zeng, "Improving sentiment classification of restaurant reviews with attention-based bi-GRU neural network," *Symmetry*, vol. 13, no. 8, p. 1517, Aug. 2021.

[10] O. Oueslati, A. I. S. Khalil, and H. Ounelli, "Sentiment analysis for helpful reviews prediction," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 7, no. 3, pp. 34–40, Jun. 2018.

[11] E. Asani, H. Vahdat-Nejad, and J. Sadri, "Restaurant recommender system based on sentiment analysis," *Mach. Learn. with Appl.*, vol. 6, Dec. 2021, Art. no. 100114.

[12] N. M. Sharef, H. M. Zin, and S. Nadali, "Overview and future opportunities of sentiment analysis approaches for big data," *J. Comput. Sci.*, vol. 12, no. 3, pp. 153–168, Mar. 2016.

[13] B. Yu, J. Zhou, Y. Zhang, and Y. Cao, "Identifying restaurant features via sentiment analysis on yelp reviews," 2017, *arXiv:1709.08698*.

[14] G. Beigi, X. Hu, R. Maciejewski, and H. Liu, "An overview of sentiment analysis in social media and its applications in disaster relief," in *Sentiment Analysis and Ontology Engineering*. 2016, pp. 313–340.

[15] O. Harfoushi, D. Hasan, and R. Obiedat, "Sentiment analysis algorithms through azure machine learning: Analysis and comparison," *Modern Appl. Sci.*, vol. 12, no. 7, p. 49, Jun. 2018.

[16] S. Gao, J. Hao, and Y. Fu, "The application and comparison of web services for sentiment analysis in tourism," in *Proc. 12th Int. Conf. Service Syst. Service Manage. (ICSSSM)*, Jun. 2015, pp. 1–6.

[17] E. Hossain, O. Sharif, M. M. Hoque, and I. H. Sarker, "SentiLSTM: A deep learning approach for sentiment analysis of restaurant reviews," 2020, *arXiv:2011.09684*.

[18] N. Hossain, M. R. Bhuiyan, Z. N. Tumpa, and S. A. Hossain, "Sentiment analysis of restaurant reviews using combined CNN-LSTM," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–5.

[19] O. Sharif, M. M. Hoque, and E. Hossain, "Sentiment analysis of Bengali texts on online restaurant reviews using multinomial Naïve Bayes," in *Proc. 1st Int. Conf. Adv. Sci., Eng. Robot. Technol. (ICASERT)*, May 2019, pp. 1–6.

[20] M. Govindarajan, "Sentiment analysis of restaurant reviews using hybrid classification method," *Int. J. Soft Comput. Artif. Intell.*, vol. 2, no. 1, pp. 17–23, 2014.

[21] O. Somantri, D. A. Kurnia, D. Sudrajat, N. Rahaningsih, O. Nurdiawan, and L. P. Wanti, "A hybrid method based on particle swarm optimization for restaurant culinary food reviews," in *Proc. 4th Int. Conf. Informat. Comput. (ICIC)*, Oct. 2019, pp. 1–5.

[22] M. K. Saad and W. M. Ashour, "Osac: Open source Arabic corpora," in *Proc. 6th ArchEng Int. Symp., EEECS*, vol. 10, 2010, pp. 1–6.

[23] O. Oueslati, E. Cambria, M. B. HajHmida, and H. Ounelli, "A review of sentiment analysis research in Arabic language," *Future Gener. Comput. Syst.*, vol. 112, pp. 408–430, Nov. 2020.

[24] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic sentiment analysis: A systematic literature review," *Appl. Comput. Intell. Soft Comput.*, vol. 2020, pp. 1–21, Jan. 2020.

[25] B. Liu, "Many facets of sentiment analysis," in *A Practical Guide to Sentiment Analysis*. Cham, Switzerland: Springer, 2017, pp. 11–39.

[26] M. Tubishat, N. Idris, and M. A. M. Abushariah, "Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges," *Inf. Process. Manage.*, vol. 54, no. 4, pp. 545–563, 2018.

[27] M. M. Agüero-Torales, M. J. Cobo, E. Herrera-Viedma, and A. G. López-Herrera, "A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor," *Proc. Comput. Sci.*, vol. 162, pp. 392–399, Jan. 2019.

[28] S. Hegde, S. Satyappanavar, and S. Setty, "Restaurant setup business analysis using yelp dataset," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 2342–2348.

[29] A. Taneja, P. Gupta, A. Garg, A. Bansal, K. P. Grewal, and A. Arora, "Social graph based location recommendation using users' behavior: By locating the best route and dining in best restaurant," in *Proc. 4th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, 2016, pp. 488–494.

[30] M. U. Khan, A. R. Javed, M. Ihsan, and U. Tariq, "A novel category detection of social media reviews in the restaurant industry," *Multimedia Syst.*, vol. 6, pp. 1–14, Oct. 2020.

[31] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment analysis in tourism: Capitalizing on big data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, Feb. 2019.

[32] A. K. A. Hassan and A. B. A. Abdulwahhab, "Reviews sentiment analysis for collaborative recommender system," *Kurdistan J. Appl. Res.*, vol. 2, no. 3, pp. 87–91, Aug. 2017.

[33] A. Kumar and A. Jaiswal, "Empirical study of Twitter and Tumblr for sentiment analysis using soft computing techniques," in *Proc. World Congr. Eng. Comput. Sci.*, vol. 1, 2017, pp. 1–5.

[34] C. Zuheros, E. Martínez-Cámara, E. Herrera-Viedma, and F. Herrera, "Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews," *Inf. Fusion*, vol. 68, pp. 22–36, Apr. 2021.

[35] M. Al Omari, M. Al-Hajj, N. Hammami, and A. Sabra, "Sentiment classifier: Logistic regression for Arabic services' reviews in Lebanon," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Apr. 2019, pp. 1–5.

[36] L. M. Alharbi and A. M. Qamar, "Arabic sentiment analysis of eateries' reviews: Qassim region case study," in *Proc. Nat. Comput. Colleges Conf. (NCCC)*, Mar. 2021, pp. 1–6.

[37] J. Qiu, C. Liu, Y. Li, and Z. Lin, "Leveraging sentiment analysis at the aspects level to predict ratings of reviews," *Inf. Sci.*, vols. 451–452, pp. 295–309, Jul. 2018.

[38] P. Pongthanoo and W. Songpan, "Feature selection and reduction based on SMOTE and information gain for sentiment mining," in *Proc. 5th Int. Conf. Comput. Commun. Syst. (ICCCS)*, May 2020, pp. 109–114.

[39] M. Scott and J. Plested, "Gan-smote: A generative adversarial network approach to synthetic minority oversampling," *Aust. J. Intell. Inf. Process. Syst.*, vol. 15, no. 2, pp. 29–35, 2019.

[40] M.-H. Nguyen, T. M. Nguyen, D. Van Thin, and N. L.-T. Nguyen, "A corpus for aspect-based sentiment analysis in Vietnamese," in *Proc. 11th Int. Conf. Knowl. Syst. Eng. (KSE)*, Oct. 2019, pp. 1–5.

[41] F. Iqbal, J. M. Hashmi, B. C. M. Fung, R. Batool, A. M. Khattak, S. Aleem, and P. C. K. Hung, "A hybrid framework for sentiment analysis using genetic algorithm based feature reduction," *IEEE Access*, vol. 7, pp. 14637–14652, 2019.

[42] A. Kumar and R. Khorwal, "Firefly algorithm for feature selection in sentiment analysis," in *Computational Intelligence in Data Mining*. Singapore: Springer, 2017, pp. 693–703.

[43] M. Tubishat, M. A. M. Abushariah, N. Idris, and I. Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1688–1707, May 2019.

[44] B. Chopard and M. Tomassini, "Particle swarm optimization," in *An Introduction to Metaheuristics for Optimization*. Cham, Switzerland: Springer, 2018, pp. 97–102.

[45] J. C. Bansal, "Particle swarm optimization," in *Evolutionary and Swarm Intelligence Algorithms*. Dhahran, Saudi Arabia: Springer, 2019, pp. 11–23.

[46] S. Sengupta, S. Basak, and R. A. Peters, II, "Particle swarm optimization: A survey of historical and recent developments with hybridization perspectives," *Mach. Learn. Knowl. Extraction*, vol. 1, no. 1, pp. 157–191, 2019.

[47] A.-Z. Ala'M, A. A. Heidari, M. Habib, H. Faris, I. Aljarah, and M. A. Hassonah, "Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems," in *Evolutionary Machine Learning Techniques*. Singapore: Springer, 2020, pp. 11–34.

[48] J. Yousif and M. Al-Risi, "Part of speech tagger for Arabic text based support vector machines: A review," *ICTACT J. Soft Comput.*, vol. 9, no. 2, pp. 1–7, Jan. 2019.

[49] A. Apsemidis and S. Psarakis, "Support vector machines: A review and applications in statistical process monitoring," *Data Anal., Appl., Comput., Classification, Financial, Stat. Stochastic Methods*, vol. 5, pp. 123–144, Apr. 2020.

[50] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: A review," *Artif. Intell. Rev.*, vol. 52, pp. 857–900, Jan. 2019.

[51] A. Gosain and S. Sardana, "Handling class imbalance problem using oversampling techniques: A review," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 79–85.

[52] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, Apr. 2018.

[53] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jul. 2018.

[54] D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance," *Inf. Sci.*, vol. 505, pp. 32–64, Dec. 2019.

[55] R. Qaddoura, H. Faris, and I. Aljarah, "An efficient clustering algorithm based on the k-nearest neighbors with an indexing ratio," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 3, pp. 675–714, Mar. 2020.

[56] X. Zheng, *SMOTE Variants for Imbalanced Binary Classification: Heart Disease Prediction*. Los Angeles, CA, USA: Univ. California, 2020.

[57] A. Alhudhaif, "A novel multi-class imbalanced EEG signals classification based on the adaptive synthetic sampling (ADASYN) approach," *PeerJ Comput. Sci.*, vol. 7, p. e523, May 2021.

[58] J. Zhang and X. Li, "Phishing detection method based on borderline-smote deep belief network," in *Proc. Int. Conf. Secur., Privacy Anonymity Comput., Commun. Storage*. Cham, Switzerland: Springer, 2017, pp. 45–53.

[59] R. M. D'Addio, M. A. Domingues, and M. G. Manzato, "Exploiting feature extraction techniques on users' reviews for movies recommendation," *J. Brazilian Comput. Soc.*, vol. 23, no. 1, pp. 1–16, Dec. 2017.

[60] A. Madasu, "A study of feature extraction techniques for sentiment analysis," 2019, *arXiv:1906.01573*.

[61] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Proc. Comput. Sci.*, vol. 152, pp. 341–348, Jan. 2019.

[62] K. Kumar, B. S. Harish, and H. K. Darshan, "Sentiment analysis on IMDb movie reviews using hybrid feature extraction method," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 5, no. 5, p. 109, 2019.

[63] J. A. García-Díaz, M. Cánovas-García, and R. Valencia-García, "Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in Latin America," *Future Gener. Comput. Syst.*, vol. 112, pp. 641–657, Nov. 2020. [Online]. Available: https://www.sciencedirect.com/3science/article/pii/S0167739X2030892X

[64] A. M. Al-Zoubi, J. Alqatawna, H. Faris, and M. A. Hassonah, "Spam profiles detection on social networks using computational intelligence methods: The effect of the lingual context," *J. Inf. Sci.*, vol. 47, no. 1, pp. 58–81, Feb. 2021.

[65] S. Srinivasan, V. Ravi, M. Alazab, S. Ketha, A.-Z. Ala'M, and S. K. Padannayil, "Spam emails detection based on distributed word embedding with deep learning," in *Machine Intelligence and Big Data Analytics for Cybersecurity Applications*. Cham, Switzerland: Springer, 2021, pp. 161–189.

[66] M. Habib, H. Faris, M. A. Hassonah, J. Alqatawna, A. F. Sheta, and A.-Z. Ala'M, "Automatic email spam detection using genetic programming with smote," in *Proc. 5th HCT Inf. Technol. Trends (ITT)*, Nov. 2018, pp. 185–190.

[67] H. Faris, J. Alqatawna, A. Z. Ala'M, and I. Aljarah, "Improving email spam detection using content based feature engineering approach," in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Oct. 2017, pp. 1–6.

[68] Y. Luo and X. Xu, "Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic," *Int. J. Hospitality Manage.*, vol. 94, Apr. 2021, Art. no. 102849.

RUBA OBIEDAT received the B.Sc. degree in computer science from The University of Jordan, in 2003, the M.Sc. degree in information system from DePaul University, in 2007, and the Ph.D. degree in e-business from the University del Salento, Lecce, Italy, in 2010. Since 2014, she has been an Associate Professor with the Department of Information Technology, The University of Jordan. Her research interests include data mining, machine learning, business intelligence, sentiment analysis, and e-business. She was awarded a full-time, competition-based Ph.D. Scholarship from the Italian Ministry of Education and Research to pursue her Ph.D. degree.

RANEEM QADDOURA received the Ph.D. degree in computer science in the fields of machine learning and data mining. She is currently an Assistant Professor at Al Hussein Technical University. She combines both academic and industrial experience of a total of 15 years. She is also an Active Research Member of the Evolutionary and Machine Learning Group (Evo-ML.com) which focuses on evolutionary algorithms, machine learning, and their applications for solving important problems in different areas. Her current research interests include evolutionary computation, deep learning, data classification, and clustering.

ALA' M. AL-ZOUBI received the B.Sc. degree in software engineering from Al-Zaytoonah University, in 2014, and the M.Sc. degree in web intelligence from The University of Jordan, Jordan, in 2017. He is currently pursuing the Ph.D. degree with the School of Science, Technology and Engineering, University of Granada. He worked as a Teacher and a Research Assistant on several projects funded by The University of Jordan. During his graduate studies, he has published several publications in well-recognized journals and conferences. His research interests include evolutionary computation, machine learning, and security in social network analysis and other research fields. He is a member of two research groups, such as the JISDF Research Group, that focus on bridge the gap between the academic and industry mechanisms in security and the Evo-ML Research Group, where the group focuses on evolutionary algorithms, machine learning, and their applications for solving important problems in different areas.

LAILA AL-QAISI received the bachelor's degree from the King Abdulla II School for Information Technology, The University of Jordan, in 2008, the master's degree in information technology management from the University of Sunderland, in 2012, and the master's degree in web intelligence from The University of Jordan, in 2017. She is currently a Lecturer at the Network Computer and Information Systems Department, Information Technology Faculty, The World Islamic Sciences and Education University. Her research interests include web and its enormous data through: artificial intelligence, machine learning, sentiment analysis, big data analytics, churn prediction, data warehouses, data mining, and cloud computing security.

OSAMA HARFOUSHI received the B.Sc. degree in CIS from the Jordan University of Science and Technology, Jordan, in 2003, the M.Sc. degree in e-business from the University of Huddersfield, U.K., and the Ph.D. degree in mobile learning from the University of Bradford, U.K. He is currently a Full Professor with the King Abdullah II School of Information Technology, Information Technology Department, The University of Jordan. His research interests include cloud computing, e-business, and business data mining.

MO'ATH ALREFAI received the B.Sc. degree in computer engineering from the Jordan University of Science and Technology, Jordan, in 2012, and the M.Sc. degree in web intelligence from The University of Jordan, Jordan, in 2017.

HOSSAM FARIS received the B.A. degree in computer science from Yarmouk University, Jordan, in 2004, the M.Sc. degree in computer science from Al-Balqa' Applied University, Jordan, in 2008, and the Ph.D. degree in e-business from the University of Salento, Italy, in 2011. In 2016, he worked as a Postdoctoral Researcher with the GeNeura Team, Information and Communication Technologies Research Center (CITIC), University of Granada, Spain. He co-founded The Evolutionary and Machine Learning (Evo-ML.com) Research Group. He is currently the Chief Data Science Officer at Altibbi. He is also a Professor with the School of Computing and Informatics, Al Hussein Technical University, and the Information Technology Department, King Abdullah II School for Information Technology, The University of Jordan, Jordan. His research interests include applied computational intelligence, evolutionary computation, knowledge systems, data mining, semantic web, and ontologies. He was awarded a Full-Time Competition-Based Scholarship from the Italian Ministry of Education and Research to pursue his Ph.D. degree in e-business at the University of Salento.

• • •