International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 2, April 2023, pp. 1891~1902
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i2.pp1891-1902

A comprehensive study of machine learning for predicting cardiovascular disease using Weka and SPSS tools

Belal Abuhaija[1], Aladeen Alloubani[2], Mohammad Almatari[3], Ghaith M. Jaradat[4], Hemn Barzan Abdalla[1], Abdallah Mohd Abualkishik[5], Mutasem Khalil Alsmadi[6]
[1]Faculty of Computer Science and Technology, Wenzhou-Kean University, Wenzhou, China
[2]Nursing Research Unit, King Hussein Cancer Center, Amman, Jordan
[3]Faculty of Sciences, Albalqa Applied University, Al-Salt, Jordan
[4]Faculty of Computer Sciences and Informatics, Amman Arab University, Amman, Jordan
[5]Faculty of Computing and Information Technology, Sohar University, Sohar, Oman
[6]Department of Management Information Systems, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Article Info

Article history:

Received Jun 24, 2022
Revised Sep 27, 2022
Accepted Oct 24, 2022

Keywords:

Attribute selection
Cardiovascular diseases
Classification
Machine learning
Statistical Package for Social Sciences
Weka

ABSTRACT

Artificial intelligence (AI) is simulating human intelligence processes by machines and software simulators to help humans in making accurate, informed, and fast decisions based on data analysis. The medical field can make use of such AI simulators because medical data records are enormous with many overlapping parameters. Using in-depth classification techniques and data analysis can be the first step in identifying and reducing the risk factors. In this research, we are evaluating a dataset of cardiovascular abnormalities affecting a group of potential patients. We aim to employ the help of AI simulators such as Weka to understand the effect of each parameter on the risk of suffering from cardiovascular disease (CVD). We are utilizing seven classes, such as baseline accuracy, naïve Bayes, *k*-nearest neighbor, decision tree, support vector machine, linear regression, and artificial neural network multilayer perceptron. The classifiers are assisted by a correlation-based filter to select the most influential attributes that may have an impact on obtaining a higher classification accuracy. Analysis of the results based on sensitivity, specificity, accuracy, and precision results from Weka and Statistical Package for Social Sciences (SPSS) is illustrated. A decision tree method (J48) demonstrated its ability to classify CVD cases with high accuracy 95.76%.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Belal Abuhaija
Faculty of Science and Technology, Wenzhou-Kean University
P88 Daxue Road, Wenzhou, China
Email: babuhaij@kean.edu

1. INTRODUCTION

The field of artificial intelligence (AI) in computer science and engineering is gaining momentum in many industries because of its ability to analyze, classify and predict future trends in finance, cyber security, image processing, and speech recognition [1]–[6]. In recent years and with the need to analyze huge data sets, AI algorithms are gaining momentum to shorten the time required to classify, analyze, and predict outcomes from medical datasets [7], [8]. In this research, we are utilizing AI algorithms to analyze, classify and predict cardiovascular diseases from a collected data set. Worldwide, cardiovascular diseases (CVDs) highest cause of mortality is caused by ischemic heart disease and stroke. It is estimated that 31% of all global deaths and more than 75% of all deaths reported in developing countries are a result of CVDs [8]. However, the

percentage increased significantly to 54% in the Eastern Mediterranean Region. Cardiovascular diseases risk factors are classified as controllable and non-controllable. The World Health Organization (WHO) considers smoking, high blood pressure, physical inactivity, obesity, diabetes mellitus, dyslipidemia, and arterial hypertension as controllable risk factors. CVDs occurrence can be significantly decreased by balancing the controllable risk factors [7], [8].

The first contribution of this paper is to utilize machine-learning algorithms in the classification and prediction CVDs on a data set. All previous research analysis conducted using machine-learning tools is limited to coronary heart disease [9]–[13] and not on the whole CVD. Therefore, the second contribution of this paper is to make a complete study of CVD, compare, and analyze the results obtained from AI algorithms and SPSS tool. The third contribution is to establish a benchmark for future studies utilizing AI and its variant predictors in gaining a better understanding of the risk factors associated with CDVs. The fourth contribution of this paper is utilizing ZeroR method as a baseline accuracy classifier.

2. LIERATURE REVIEW

Several systematic reviews are presented in [14]–[19] where the mortality rate in smoking individuals increases by almost three folds. Many studies have investigated the cardiovascular diseases (CVD) risk factors among university students and staff [20], [21]. A study conducted among students in Colombia showed that 92% of students had a low risk of cardiovascular disease, 2% and 6%, more than 50% of the students and staff had at least one risk factor, including overweight or obesity, sedentary lifestyle, and hypertension [14], [17].

A cross-sectional study was conducted in Turkey to identify university students' awareness of cardiovascular risk factors (n=2450). The students perceived smoking (58.7%), stress (71.8%), high cholesterol (72.3%), obesity (64.3%), diabetes (52.7%), inactivity (47.8%), hypertension (64.2%), and a family history of CVD (44.4%) to be the main risk factors for CVD. Moreover, the results showed that men ignore these risk factors [21]. There are limited studies aimed at identifying the prevalence and awareness of risk factors for CVDs among university staff and students, especially in the Middle East [20], [21]. In addition, a systematic review of 212 articles aimed to identify models for predicting risk factors for CVDs in the general population [19]. Most prediction studies in this review were carried out in Europe (n=167, 46%) for both fatal and non-fatal coronary heart disease (n=118, 33%) and for more than ten years (n=209, 58%). Smoking, age, gender, blood pressure, blood cholesterol, diabetes, body mass index, and hypertension were the most frequent predictors [19]–[22].

Weka is applied to a number of collected datasets in the past two decades for cardiovascular diseases including heart disease. Four well-known datasets dedicated to heart disease were introduced in 1988 and provided by the UCI machine learning repository [10]. Many research works in the literature presented and implemented a number of machine learning methods to predict the presence of a heart attack among patients from four different medical institutes. Some examples of implementing machine-learning methods to these particular datasets are presented by Rao [23], who applied a variety of classifiers including k-nearest neighbor (k-NN) with poor accuracy and random forest (RF) with the highest accuracy of around 90%. Some other implementations were directed toward different datasets other than the heart disease dataset by the UCI repository applied a convolutional neural network (CNN) which obtained a good accuracy of 82% [24]; Haq *et al.* [9] applied a variety of machine learning algorithms. Lopez-Martinez *et al.* [10], [11] applied conventional artificial neural networks (ANN). Uyar and İlhan [13] applied genetic algorithms and recurrent fuzzy neural networks. Others focused on handling overfitting the predictive model through optimization metaheuristics such as [25]. Several studies are similar to our research approach but with different datasets (e.g., ultrasound images, UCI, and Kaggle numerical data) and machine learning methods [26]–[30]. Another similar study which utilized Weka with highly accurate predictions of a UCI diabetes dataset is presented by Alalwan [31].

This research intends to produce and analyze a complete study of the most crucial parameters in CVD and utilize machine-learning methods for classification tasks to predict the presence of heart disease. Machine learning methods are very useful in identifying hidden patterns and information in the dataset. Weka and Statistical Package for Social Sciences (SPSS) tools has several classifiers such as baseline accuracy, decision tree (DT), naïve Bayes (NB), *k*-NN, support vector machines (SVM), and ANN (including multilayer perceptron and backpropagation). A comparison of the results based on sensitivity, specificity, precision, area under curve, and accuracy is realized. We differ from [10] by utilizing ZeroR method the baseline accuracy classifier as a performance benchmark of other classifiers which makes our predictive models' fitness and efficiency more accurate than what has been presented by Ramotra *et al.* [10]. Furthermore, the authors, did not report their model's configuration clearly, might have triggered training and testing using carefully tuned parameters settings.

We have utilized default configurations and parameter settings for each model in both Weka and SPSS-modeler tools to conduct a clear and fair comparison. In addition, we have implemented a multi-layer perceptron (MLP) neural network with a backpropagation strategy, which gave us a better understanding of the dataset's parameters and information as well as the best-suited predictive model to be performed for heart disease. Our results show that the SVM classifier achieved the highest accuracy of 95.94% in the Weka tool, and in the SPSS-Modeler tool closely followed by J48, MLP, and multi-layer perceptrons with back propagation (MLPBP) classifiers. Although SPSS-modeler has slightly outperformed Weka in terms of accuracy (e.g., *k*-NN, linear regression (LR), SVM), however, it clearly shows overfitting models based on sensitivity, area under curve, the variance between sensitivity and precision. Hence, Weka has demonstrated that it is much robust in modelling predications than SPSS-modeler.

3. METHODS

This section presents the design methodology of seven classifiers (predictors). It illustrates well known and popular classifiers including baseline accuracy (ZeroR), NB, *k*-NN, DT (J48), SVM, LR, and ANN MLP. Their architectures resulting in the processing of datasets, as well as the implementation details of the classification process, are also discussed.

3.1. Classification and Weka Workbench

Classification is considered as a supervised learning system that uses a labelled dataset representing predictions. The dataset is used as a training set of input-output pairs to realize a deterministic function that maps inputs to outputs. Then predicting future input-output observations while minimizing errors as much as possible. ANNs, including MLP, are one of the best techniques for classifying data organized in a tabular form or a warehouse. Weka is a free workbench software that holds a collection of algorithms and visualization tools for data analysis and prediction models. It comes with an easy-to-use graphical interface for all of its functionalities. In this study, we are utilizing Weka to perform the classification and clustering tasks by implementing the above mentioned seven algorithms. Weka has a variety of settings with many parameters tuning functions. This gives the authors a rich environment of experiments with less time.

3.2. The implemented models

A machine learning model must be capable of recognizing patterns in a dataset which looks for certain features of a, let us say, heart disease case such as blood pressure or cholesterol [32], [33]. To overcome the issue of detecting a heart disease case across a huge number of medical data, in other words generalizing the model, it is required to learn e.g., ANN to detect certain features. Thus, a machine learning model will filter a group of data for extracting significant features (or attributes) mainly ranking their significance via principal component analysis and attribute subset evaluator filters or wrappers. Then those selected features are then proceeded for further classification processing that results in prediction.

The implemented predictive models in our study are:

a. ZeroR: a rules-based classifier that is the simplest classification method which relies on the target and ignores all predictors. It simply predicts the majority class which is only useful for determining a baseline performance as a benchmark for other classification methods. A very limited number of research papers have considered baseline accuracy classifiers in their work.
b. NB: naïve Bayes classifier which is a collection of classification algorithms based on Bayes' theorem, where all of them share a common principle, e.g., every pair of features being classified is independent of each other.
c. *k*-NN: *k*-nearest neighbor lazy-based classifier that is one of the simplest decision procedures. It classifies a sample based on the class of its nearest neighbor. For large samples, it has a twice less probability of error compared to any other decision rule. It uses some or all the patterns available in the training set to classify a test pattern. It involves finding the similarity between the test pattern and every pattern in the training set.
d. LR: logistic regression function-based classifier. It is used to come up with a hyperplane in feature space to separate observations that belong to a class from all other observations that do not belong to that class. The decision boundary is thus linear.
e. J48: An optimized Java-based version of the C4.5 DT-based classifier. The C4.5 algorithm is used to generate a decision, based on a certain sample of data (univariate or multivariate predictors) using the concept of information entropy.
f. SVM: support vector machine function-based classifier. It is able to generalize between two different classes if the set of labelled data is provided in the training set to the algorithm. The main function of the SVM is to check that hyperplane is able to distinguish between the two classes.

g. MLP: multilayer perceptron function-based classifier. MLP is a class of feedforward ANN where it has a sophisticated architecture, consisting of 3 layers of nodes including input, hidden, and output layers. It utilizes a non-linear activation function and backpropagation for supervised training processes to define the relations between inputs and outputs for more flexibility. In brief, the MLP is:
   - Forward pass (input layer): passing inputs into the model multiplied by weights and adding bias at every layer to discover the model calculated output.
   - Loss calculates (activation function/hidden layer): once a sample data (e.g., a record in a dataset) an output is obtained from the model as a predicted output, which will be labelled with the data that is a real or expected output. Hence, embedded backpropagation algorithm is used to calculate the loss.
   - Backward pass (output layer): finally, and most importantly in training the model, the loss is back propagated and by using gradient updates the weights of the model. Iteratively, weights will be adjusted referring to the gradient flow to a certain direction.

Based on the presented characteristics of the algorithms, we are motivated to adopt a machine-learning model that utilizes a more sophisticated classification architecture including training and testing such as MLP. Its architecture presents a robust and effective one, where it provides a strong gradient flow due to implicit supervision (e.g., learning), parameters, and computational efficiency due to strong connectivity through a number of parameters which is directly proportional to the growth rate of the complexity, more diversified features, and maintains low complexity features. Therefore, in this study, we employ 7 traditional well-known machine learning algorithms including MLP for 1 class attribute of heart disease related to cardiovascular classification, training, validation, and testing. We train the model using the dataset collected by the authors alongside selecting significant features. The parameter values that give rise to the best performance on the validation dataset are used for testing.

4. COMPUTATIONAL RESULTS AND DISCUSSION

This section discusses the bases of training and testing the predictive models including their experimental settings and results. Datasets that are used in this work are also briefly discussed. Using a Java-based workbench called Weka version 3.9.4, experiments are conducted on a Windows 10 machine with a Corei7 processor assessed by Nvidia GPU and 16 GB of RAM. The following subsections demonstrate the training and testing environment, parameter settings of the predictive model, and configuration of machine learning algorithms used within the predictive model.

4.1. The implemented models

The data was collected from students, employees, and faculty members of a Saudi University; data from 370 participants (240 males and 130 females), including university employees and students, in Saudi Arabia were analyzed. The majority of participants were Saudi Arabian (82.2%) aged 18-25, unemployed (79.5%), and married (54.6%). We are utilizing the Weka machine-learning tool to conduct this study on a data set from some university students, faculty, and employees in the middle east. In order to validate our approach, Weka and the SPSS tool findings and results are shown below. A number of consecutive steps are considered in conducting the experiments.

a. Data collection is to collect relevant data, define and understand the problem, then are formatted into a Weka-based arff dataset, where the class attribute is labeled for the classification task. The dataset has 370 records and 36 attributes including 1 class attribute. In general, the quantity and quality of the dataset dictate how accurate the model is. The outcome of this step is a representation of data in a tabular form which will be used for training the model.
b. Data preparation is the step of wrangling data and preparing it for training. It includes filtering data and cleaning it via removing duplicates, correcting errors, dealing with missing values, normalization, and data type conversion.
c. Feature selection is selecting the impactful subset of attributes on the model performance and its accuracy.
d. Choosing a model that fits the classification and regression tasks.
e. Training the model to iteratively predict correctly as often as possible.
f. Evaluate the model using a combination of metrics to measure the objective performance of the model. In addition, testing the model against previously unseen data, which mimics the model performance in the real world and to tune the model furthermore.
g. Parameter tuning (a.k.a. hyperparameter tuning) which may lead to an improved performance of the model. For example, number of training steps, learning rate, initialization values, and distribution.
h. Perform predictions using further test dataset that is new to the model but for which class labels are known to test the model and obtain a better approximation of the performance of the model in the real world.

We have considered a comparative experiment between three levels of analysis including descriptive statistics using IBM-SPSS, then classification and regression using machine-learning techniques in Weka workbench, and finally predictive modeling using a machine learning technique. IBM-SPSS is used to understand and statistically describe the collected data by identifying central tendency measures, coefficient correlations, and the significance of attributes. Using Chi-square, for instance, we have identified some significant attributes that may affect the predictive model's accuracy. However, this did not sufficiently determine the significance and correlation coefficient of attributes, hence we further experiment with those measures using Weka workbench for both levels of training models machine learning. Machine learning consists of a variety of induction and statistical based functions that describe the data via sophisticated techniques e.g., SVM and MLP which help to learn processes towards predictions.

5. EXPERIMENTAL DATASETS

The datasets that are used in this research are collected by a questionnaire, where the whole dataset (size=370 instances) is split into two sets, training, and testing in the form of "arff" Weka file which consists of 36 attributes including 7 class attributes. Train set is used to train a model for prediction, while the test set is used to feed those models with new data for accuracy evaluation. Experimentally predetermined, the training set is 70% split while the test set is 30% split. Table 1 summarizes labels and their count for each class attribute. Compared to the well-known UCI heart disease datasets, they have 76 attributes that are reduced into 14 significant attributes pointed to only one class (e.g., heart disease), while our dataset has 36 attributes that are reduced into a handful of significant attributes (2 to 5 attributes) pointed to 7 classes not only for the heart disease diagnosis. It is easier, faster, and more accurate to predict a class using a small set of attributes. With the 7 classes in our dataset, we are able to diagnose 7 different (but related to heart disease) CVDs. In addition, our dataset has no missing values as in the UCI heart disease datasets, which makes it easier to process and better to understand. Some significant attributes are common between our dataset and the UCI datasets such as age, gender, cholesterol, fasting blood sugar, and blood pressure. Those contribute to our dataset's validity. The dataset has the float numerical readings of cardiovascular indexes and related diseases, e.g., fasting blood sugar, low and high blood lipids, triglyceride, cholesterol, body mass index, blood pressure, demographics, and some smoking and eating habits [34], [35]. The readings present some distinct features of heart disease cases that are clearly associated with other features such as blood pressure and fasting blood sugar that are well known to medical experts and therapists.

Table 1. Dataset class attribute and labels

Class	Label	Count
Heart Disease (HD)	Yes	112
	No	258

6. EXPERIMENTAL RESULTS

Here, we summarize obtained accuracy results of machine learning predictive models. For the classification task we have implemented seven popular machine learning algorithms with the same experimental settings and configuration, e.g., data preprocessing, testing mode, and attribute selection method. The parameter settings of the seven models with many common parameters and setups which are predetermined experimentally. For example, the 10 folds cross-validation test mode. The following justification illustrates the reason behind applying such a testing mode "cross-validation" as a suitable mode for our experiments.

Generally, MLP overcomes the issue of traditional machine learning methods in processing data, where it uses perceptrons for input. Using filters or wrappers within MLP, an ANN analyzes the influence of correlated coefficients (e.g., selected attributes) in a dataset that is more powerfully associated than others, and subsets are built upon instances. After passing a number of filters for the whole dataset, then on each filter, a feature subset is generated. These feature subsets are then passed over an activation function to choose whether a certain feature is existent at a given position (or rather a correlation) in the data. In short, MLP is capable of learning the following: i) each layer learns filters of growing complexity; ii) first layers learn basic feature detection filters e.g., correlation, and significance; iii) middle layers learn filters that detect subsets of attributes; and iv) last layers have higher demonstrations by learning to identify the best evaluated and significant subset of attributes (features), in diverse measurements.

One of the predictive model evaluation techniques is the cross-validation technique which splits the dataset into *k*-folds. It is training a model on all of the folds except one that is held out as the test set, then this process repeatedly creates *k*-different models and give each fold a chance of being held out as the test set.

Then calculate the average performance of all *k* models. Commonly used, this is the gold standard for evaluating model performance but has the cost of creating many more models. The training dataset technique is only used when having the whole dataset and needs to create a descriptive model rather than the predictive model. This way, we will be able to create a model to better understand the problem. This was achieved by using the IBM-SPSS software. The supplied test set is only used when having a very large dataset which is not applicable in our experiment where the dataset size=370 instances. The percentage split technique is used when needing a quick idea of the model performance and could be impractical for making decisions. Hence, cross-validation is the default option to be used when uncertainty about the problem's description. Generally, it provides a more accurate estimate of the performance than the other evaluation techniques, not to be used when having very large data. Common values for *k* are 5 and 10, depending on the size of the dataset. In short, cross-validation is a popular technique because it is simple to understand and it generally results in a less biased or less optimistic estimate of the model performance than others, such as a simple train/test split. Tables 2 to 5 present the accuracy outcomes of classification models using Weka and SPSS compared against each other.

Table 2. Baseline classifier employed in Weka and SPSS

ZeroR	Sensitivity	Specificity	Precision	Area under curve	Accuracy (%)
SPSS	0.000	1.000	0.000	0.834	69.70%
Weka	0.697	0.000	0.697	0.940	69.72%

Table 3. Naïve Bayes classifier employed in Weka and SPSS

NB	Sensitivity	Specificity	Precision	Area under curve	Accuracy (%)
SPSS	0.000	1.000	0.000	0.916	72.63%
Weka	0.943	0.973	0.945	0.985	94.32%

Table 4. *k*-NN classifier employed in Weka and SPSS

k-NN	Sensitivity	Specificity	Precision	Area under curve	Accuracy (%)
SPSS	0.604	0.969	0.878	0.917	85.78%
Weka	0.854	0.517	0.854	0.842	85.40%

Table 5. Logistic regression classifier employed in Weka and SPSS

LR	Sensitivity	Specificity	Precision	Area under curve	Accuracy (%)
SPSS	0.000	1.000	0.000	0.974	93.68%
Weka	0.922	0.892	0.923	0.954	92.16%

Determining the baseline accuracy of each class attribute using ZeroR rule-based machine learning algorithm is needed at first to benchmark the performance of a predictive model. This is to verify how good the accuracy percentage obtained by the predictive models is. ZeroR typically guesses the most popular class attribute, e.g., by taking 258 instances from the heart disease (HD) class attribute and dividing it into 370 the whole instance's size which equals to 69.72% accuracy. This is a good indicator that e.g., *k*-NN=85.4% or J48=95.67% are highly accurate compared to the baseline accuracy (69.72%), but the accuracy is not a sufficient indicator of a highly accurate model that is suitable for the data. Overfitting or underfitting the model is validated using measurements such as precision, sensitivity, and specificity. Therefore, understanding the data and its pattern is crucial and supported by the baseline classifier (ZeroR) before preprocessing data or fine-tuning a model's parameters. Refer to Tables 2 to 5. We have implemented default parameter settings of the predictive models including training and testing mode. Based on the confusion matrix obtained by all classifiers, some crucial measures and rates are described:

- Accuracy (correctly classified instances) presents the percentage of test instances that were correctly and incorrectly classified. However, it is not chance corrected and not sensitive to class distribution. So, it is an insufficient performance measurement.
- True positive rate (TP) presents when the model actually predicts "Yes", how often does it predict "Yes"? For example, TP/actual "Yes" =100/112=0.89. In other words, instances are correctly classified as a given class.
- False positive rate (FP) presents when the model actually predicts "Yes", how often does it predict "No"? For example, FP/actual "Yes" =12/112=0.11. In other words, instances that are incorrectly classified as a given class.

- Precision presents when the model predicts "yes", how often is it correct? For example, where a proportion of instances that are true of a class is divided by the total instances classified as that class; TP/predicted "No" =258/270=0.95.
- Sensitivity (aka. recall) is a proportion of instances classified as a given class divided by the actual total in that class which is equivalent to TP rate (TP/(TP+FN)). It is used when the occurrence of false negatives is unacceptable. It would be better to have extra false positives over some false negatives. It is very useful when predicting disease.
- Specificity is a ratio of true negatives to total negatives in the data (TN/(TN+FP)). It is used when stopping false alarms is required. It is very useful when running e.g., a blood pressure test in which all patients who test positive will immediately be classified as heart attack potentials.

ZeroR classifier is not provided explicitly in SPSS and SPSS-modeler, which need a number of cumulative steps, while it is fully provided in Weka with full control of the classifier's parameters that are easy to tune in one workspace. So, we are unable to report the outcomes of the baseline classifier in a fair and comparable form. As mentioned before, baseline accuracy (as a performance benchmark evaluation) is crucial for testing and validating a classification process or predictive model. This favors Weka over SPSS due to automation, benchmarking, and parameter control. From Table 2, both software showed very similar performance with an almost identical accuracy percentage. However, Weka has obtained better indicators of how to fit the classifier model to the collected dataset. See both models' sensitivity, specificity, precision, and area under the curve. Note that SPSS performs a default 50/50 split of the dataset for the baseline accuracy, hence the performance benchmark is not worthy.

Again, we are unable to report the outcomes of the NB classifier from SPSS in a fair and comparable form since SPSS does not provide full control of NB classifier's parameters. Most importantly, NB classifier of SPSS only reports the distribution of the Bayesian network in which it translates the conditional probabilities of the heart disease class. It distributed the probability of the categories of the class into "Yes"=0.27 and "No"=0.73. From Table 3, it is shown that NB classifier from Weka has outperformed the one from SPSS in terms of area under the curve as well as a very high accuracy percentage with a big difference.

From Table 4, both Weka and SPSS are almost identical in terms of accuracy. However, observing the relation between sensitivity and area under the curve, SPSS has clearly built an overfit model. This is due to the fact default configuration is used in *k*-NN model, however, in the SPSS dataset split for training and testing. Hence, Weka has built a fit predictive model based on the observation of sensitivity. Table 5 shows Weka and SPSS are approximately equal in terms of accuracy, however, the area under the curve is high, and the variance between both software in sensitivity and precision indicates the limitation of SPSS, which renders Weka's superiority to SPSS.

Figure 1 shows that Weka has outperformed SPSS-modeler in terms of both accuracy and model fit. Therefore, it is evident that Weka is more suitable than SPSS as it can overcome the limitations of SPSS. As far as the SVM algorithm is concerned, Figure 2 illustrates the results obtained from Weka and SPSS. SPSS has obtained a better accuracy percentage and higher area under the curve, but at a great cost giving away suitability for the dataset by overfitting the model. Hence, Weka showed a model that fits the dataset.

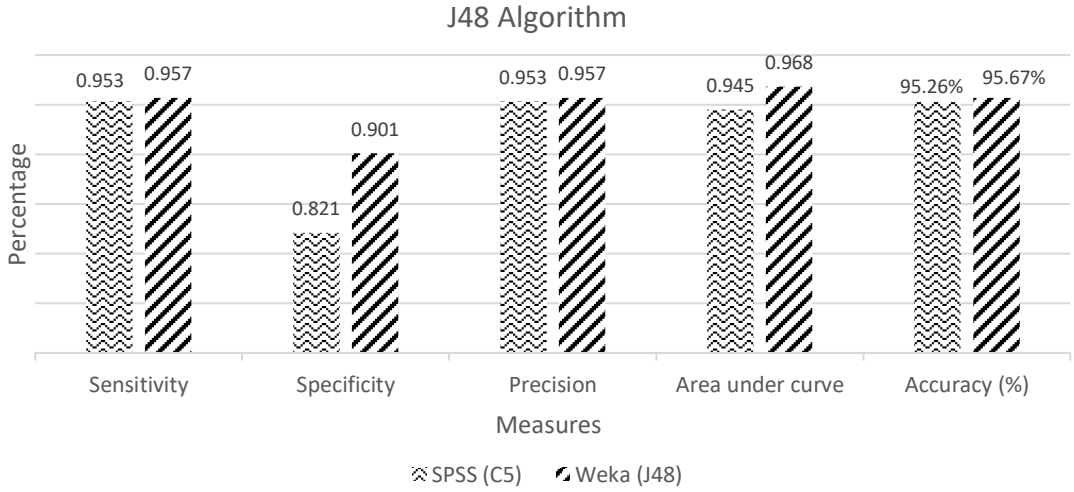

Figure 1. Decision tree classifier employed in Weka and SPSS

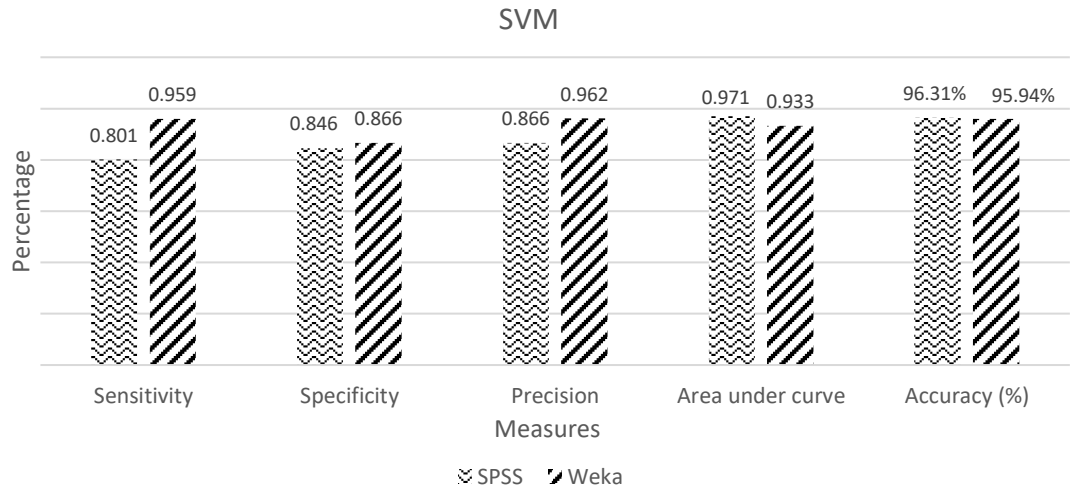

Figure 2. SVM classifier employed in Weka and SPSS

MLP using Weka has obtained better accuracy (94.59%) than the one using SPSS (87.36%). Furthermore, Figure 3 illustrates a better accuracy and avoided overfitting presented by a higher area under curve in relation to sensitivity. Figure 4 illustrates that only Weka provides MLPBP explicitly, therefore, there were no results from SPSS. MLPBP has obtained the best overall performance among all classifiers presented in Tables 2 to 5 and Figures 1 to 4.

With a total number of 370 instances in the dataset, it is clear from Tables 1 to 5 and Figures 1 to 4 that all classifiers have obtained highly accurate classifications. It is clear that J48, MLP, and MLPBP models have obtained the highest accuracy compared to the baseline accuracy for all classes, some of which are excellent (e.g., SVM) or acceptable while others are poor (e.g., *k*-NN). Some models are overfitting such as J48, NB, LR, and MLP, while others are underfitting models such as *k*-NN and SVM. MLPBP model has reached stability without overfitting nor underfitting the data.

The best experimental configuration of the best performing MLPBP model is: number of layers is 3, number of inputs is 35, number of hidden layers is 18 *((Attributes+Classes)/2=18)*, number of outputs is 2, learning rate of 0.3, momentum of 0.2 to update weights, 100 epochs and cross-validation training as a testing mode, recursive instance iterator, 500 seconds running time, *Softplus* as an activation function, *SquaredError* as a loss function, a steepest descent heuristic as an optimization algorithm, and a subset attribute selector with best first search within attributes space.

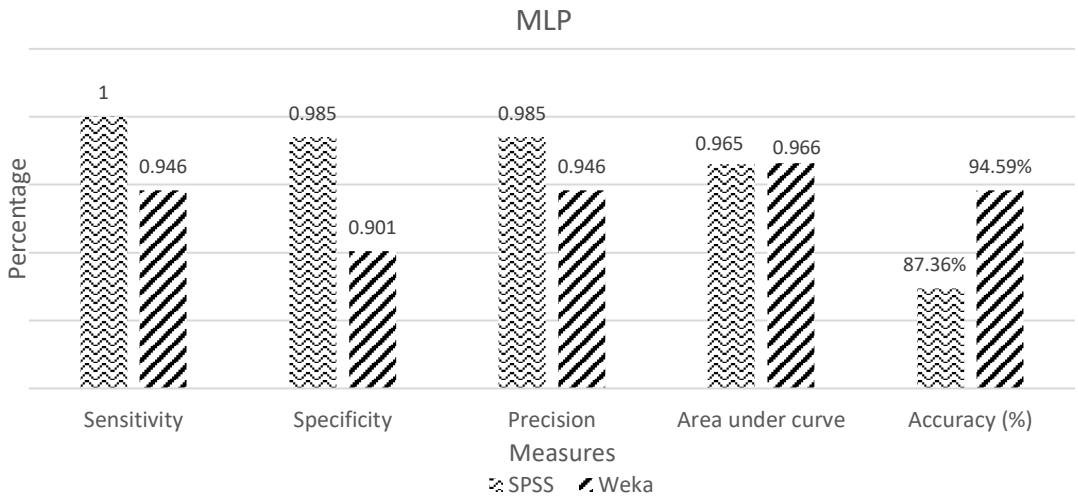

Figure 3. MLP classifier employed in Weka and SPSS

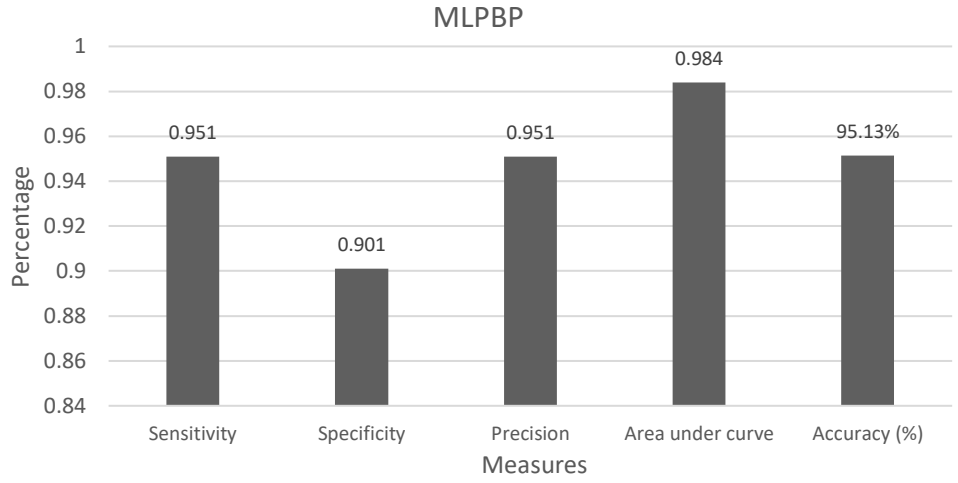

Figure 4. MLP-backpropagation classifier employed in Weka and SPSS

7. CONCLUSION

The results indicate a clear impact on the classification accuracy, which can be contributed to a number of characteristics in the MLPBP model. The first characteristic is the number of hidden layers in relation to the testing mode (number of attributes plus number of classes). The second is the cross-validation testing mode that proved better than both percentages split and training evaluation modes for all class attributes. This could be referred to as the size of the training dataset. This observation also applies to the rest of the models. Finally, the third is the optimization algorithm that employed the steepest descent heuristic, which has the ability to converge quickly towards local minima. It passes the best-fitted information back and forth between the input and output layers.

These characteristics are subject to find out attributes that have the greatest impact on the classification task. In order to obtain accurate results for the predictive models, we need to select relatively large, independent features that can result in better validation. Our MLPBP model obtained almost all the predictions by integrating medical observations and demographic data. However, the choice of the most appropriate algorithm depends on many parameters including the types of data collected, the size of the data samples, the time limitations as well as the type of prediction outcomes. The MLPBP model has proved to be accurate, rapid, and inexpensive medical decision making. It has handled test datasets robustly and efficiently. It also obtained highly accurate classifications compared to other classifiers including the baseline classifier.

SPSS failed to outperform Weka due to using the Chi-square test, although it is very handy for fast training, its accuracy is medium due to its classification limits and the inability to differentiate correlated predictors. Chi-square is very useful to determine statistically significant attributes only in the demographic part of our dataset. That is why we implemented Chi-square using IBM-SPSS prior to deploying Weka. Weka has proved to be better than SPSS in different aspects such as the use of baseline benchmark, full control of classifiers' parameter settings, default configurations, full access of performance metrics and model's fitness, providing a wider range of classifiers and modelers, a wider variety of data filters, wrappers, and features selection methods, and finally automation.

In future research, we will make more efforts trying to improve the accuracy of the predictive model and enhance its performance, including parameter tuning. The improvement will focus on implementing two main modifications. First, employing a better feature selection technique to select highly correlated attributes to their respective class (e.g., heart disease). This can be achieved by using wrapping subsets and evaluating them and a powerful optimization algorithm, then employing a better learning strategy to control bias, convergence, and avoiding overfitting. We can achieve this by applying a SoftMax activation function, sigmoid updater, and non-improvement every 5 epochs.

REFERENCES

[1] M. K. Alsmadi, "Fish recognition based on robust features extraction from size and shape measurements using neural network," *Journal of Computer Science*, vol. 6, no. 10, pp. 1088–1094, Oct. 2010, doi: 10.3844/jcssp.2010.1088.1094.

[2] M. K. Alsmadi, "Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 3, pp. 373–381, Jul. 2018, doi: 10.1016/j.jksuci.2017.05.002.

[3] M. K. Alsmadi and I. Almarashdeh, "A survey on fish classification techniques," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 1625–1638, May 2022, doi: 10.1016/j.jksuci.2020.07.005.
[4] R. M. A. Mohammad and M. K. Alsmadi, "Intrusion detection using highest wins feature selection algorithm," *Neural Computing and Applications*, vol. 33, no. 16, pp. 9805–9816, Aug. 2021, doi: 10.1007/s00521-021-05745-w.
[5] B. A. Aldeeb *et al.*, "Hybrid intelligent water drops algorithm for examination timetabling problem," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 4847–4859, Sep. 2022, doi: 10.1016/j.jksuci.2021.06.016.
[6] D.-J. Lee, R. B. Schoenberger, D. Shiozawa, X. Xu, and P. Zhan, "Contour matching for a fish recognition and migration monitoring system," in *Two-and Three-Dimensional Vision Systems for Inspection, Control, and Metrology II*, 2004, pp. 37–48.
[7] WHO, "Cardiovascular diseases (CVDs)," *World Health Organization*, 2021. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Jun. 13, 2021).
[8] WHO, "Cardiovascular diseases," *World Health Organization*. https://www.emro.who.int/health-topics/cardiovascular-diseases/introduction.html (accessed Jun. 13, 2021).
[9] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Information Systems*, pp. 1–21, Dec. 2018, doi: 10.1155/2018/3860146.
[10] A. K. Ramotra, A. Mahajan, R. Kumar, and V. Mansotra, "Comparative analysis of data mining classification techniques for prediction of heart disease using the Weka and SPSS modeler tools," in *Smart Innovation, Systems and Technologies*, Springer Singapore, 2020, pp. 89–97.
[11] F. Lopez-Martinez, A. Schwarcz, E. R. Nunez-Valdez, and V. Garcia-Diaz, "Machine learning classification analysis for a hypertensive population as a function of several risk factors," *Expert Systems with Applications*, vol. 110, pp. 206–215, Nov. 2018, doi: 10.1016/j.eswa.2018.06.006.
[12] R. Nakanishi *et al.*, "Machine learning in predicting coronary heart disease and cardiovascular disease events: results from the multi-ethnic study of atherosclerosis (MESA)," *Journal of the American College of Cardiology*, vol. 71, no. 11, Mar. 2018, doi: 10.1016/S0735-1097(18)32024-2.
[13] K. Uyar and A. İlhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," *Procedia Computer Science*, vol. 120, pp. 588–593, 2017, doi: 10.1016/j.procs.2017.11.283.
[14] V. Colpani *et al.*, "Lifestyle factors, cardiovascular disease and all-cause mortality in middle-aged and elderly women: a systematic review and meta-analysis," *European Journal of Epidemiology*, vol. 33, no. 9, pp. 831–845, Sep. 2018, doi: 10.1007/s10654-018-0374-z.
[15] S. Irawati *et al.*, "Long-term incidence and risk factors of cardiovascular events in Asian populations: systematic review and meta-analysis of population-based cohort studies," *Current Medical Research and Opinion*, vol. 35, no. 2, pp. 291–299, Feb. 2019, doi: 10.1080/03007995.2018.1491149.
[16] N. A. M. Nor, A. Mohamed, and S. Mutalib, "Prevalence of hypertension: predictive analytics review," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, pp. 576–583, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp576-583.
[17] D. M. B. Enriquez, K. V. L. Oyola, and J. E. Y. Ruiz, "Cardiovascular risk factors and diabetes in medical students: observational study, experience in Colombia," *International Journal of Medical Students*, vol. 6, no. 2, pp. 61–65, Jul. 2018, doi: 10.5195/ijms.2018.24.
[18] D. Boateng *et al.*, "Knowledge and awareness of and perception towards cardiovascular disease risk in sub-Saharan Africa: A systematic review," *Plos One*, vol. 12, no. 12, Dec. 2017, doi: 10.1371/journal.pone.0189264.
[19] J. A. A. G. Damen *et al.*, "Prediction models for cardiovascular disease risk in the general population: systematic review," *BMJ*, May 2016, doi: 10.1136/bmj.i2416.
[20] H. Hsieh, "The relationship between physical activity behavior and cardiovascular health among university employees," Master Thesis, Clinical Exercise Physiology, UWRF, 2018.
[21] F. E. Güneş, N. Bekiroglu, N. Imeryuz, and M. Agirbasli, "Awareness of cardiovascular risk factors among university students in Turkey," *Primary Health Care Research and Development*, vol. 20, Sep. 2019, doi: 10.1017/S146342361900063X.
[22] CDC, "Diabetes," *Centers for Disease Control and Prevention*. https://www.cdc.gov/diabetes/basics/getting-tested.html (accessed Nov. 08, 2020).
[23] J. N. Rao, "Cardiovascular disease prediction using machine learning techniques," *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, pp. 6875–6880.
[24] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An efficient convolutional neural network for coronary heart disease prediction," *Expert Systems with Applications*, vol. 159, Nov. 2020, doi: 10.1016/j.eswa.2020.113408.
[25] Z. Yang, T. Zhang, J. Lu, D. Zhang, and D. Kalui, "Optimizing area under the ROC curve via extreme learning machines," *Knowledge-Based Systems*, vol. 130, pp. 74–89, Aug. 2017, doi: 10.1016/j.knosys.2017.05.013.
[26] J. O. (Ryan) Kim, Y.-S. Jeong, J. H. Kim, J.-W. Lee, D. Park, and H.-S. Kim, "Machine learning-based cardiovascular disease prediction model: a cohort study on the Korean National health insurance service health screening database," *Diagnostics*, vol. 11, no. 6, May 2021, doi: 10.3390/diagnostics11060943.
[27] S. Beguma, F. A. Siddiqueb, and R. Tiwari, "A study for predicting heart disease using machine learning," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, pp. 4584–4592, 2021.
[28] S. Shareefunnisa, D. R. Mahima, and Y. Padma, "Cardiovascular heart disease prediction using machine learning classifiers with data mining techniques," *Turkish Online Journal of Qualitative Inquiry*, vol. 12, no. 3, 2021.
[29] N. Sureja, B. Chawda, and A. Vasant, "A novel salp swarm clustering algorithm for prediction of the heart diseases," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 25, no. 1, pp. 265–272, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp265-272.
[30] H. Sofian, J. T. C. Ming, S. Muhammad, and N. Mohd Noor, "Calcification detection using convolutional neural network architectures in Intravascular ultrasound images," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 17, no. 3, pp. 1313–1321, Mar. 2020, doi: 10.11591/ijeecs.v17.i3.pp1313-1321.
[31] S. A. D. Alalwan, "Diabetic analytics: proposed conceptual data mining approaches in type 2 diabetes dataset," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 14, no. 1, pp. 88–95, Apr. 2019, doi: 10.11591/ijeecs.v14.i1.pp88-95.
[32] C. Li, G. Engström, B. Hedblad, G. Berglund, and L. Janzon, "Blood pressure control and risk of stroke," *Stroke*, vol. 36, no. 4, pp. 725–730, Apr. 2005, doi: 10.1161/01.str.0000158925.12740.87.
[33] E. R. Atkins and V. Perkovic, "Blood pressure: at what level is treatment worthwhile?," *Australian Prescriber*, vol. 42, no. 4, Aug. 2019, doi: 10.18773/austprescr.2019.038.
[34] R. B. Jain and A. Ducatman, "Associations between smoking and lipid/lipoprotein concentrations among US adults aged ≥20 years," *Journal of Circulating Biomarkers*, vol. 7, Jan. 2018, doi: 10.1177/1849454418779310.

[35] A. L. Merianos, R. A. Jandarov, J. C. Khoury, and E. M. Mahabee-Gittens, "Tobacco smoke exposure association with lipid profiles and adiposity among U.S. adolescents," *Journal of Adolescent Health*, vol. 62, no. 4, pp. 463–470, Apr. 2018, doi: 10.1016/j.jadohealth.2017.10.001.

BIOGRAPHIES OF AUTHORS

Belal Abuhaija is an associate/assistant professor at the Faculty of Computer Science and Technology at Wenzhou-Kean University, China. Belal has finished his bachelor's and master's degree from the USA in electrical engineering and telecommunications networks respectively and his Ph.D. from the UK in cellular and wireless networks. Belal has many industrial certificates such as CCNP, CCNA Instructor, and CISCO Cyber Security operations. He has been certified by the British Computer Society (BCS) as a trainer for blockchain and information security. His research interests are in various fields, such as 4G/5G networks security and optimization as well as artificial intelligence and blockchain technology. He can be contacted at babuhaij@kean.edu.

Aladeen Alloubani completed his bachelor's degree in nursing in 2007 and his M.Sc. and Ph.D. degrees in health management in 2009 and 2015, respectively. In 2019, he completed a post-graduate specialty program in High Impact Cancer Research at Harvard Medical School in the United States. From 2007 to 2010, Aladeen worked in the intensive care unit at Istishari Hospital. Between 2010 and 2017, he worked as a lecturer at the Faculty of Nursing at the University of Tabuk. In 2017, he joined King Hussein Cancer Center as a senior manager for nursing research and evidence-based practice. Aladeen now supervises all nursing research and evidence-based practice (EBP) activities, including quantitative and qualitative designs. Moreover, he is conducting a one-year education program titled "nursing research and EBP" and organizing international conferences in collaboration with other departments. He can be contacted at aa.12567@khcc.jo.

Mohammad Almatari is an assistant professor in the Department of Physics in the Faculty of Sciences at Al-Balqa Applied University in Jordan since 2020. From 2012 to 2020 he has been working as an assistant professor at Tabuk University. He received a B.Sc. from Applied Science University in 2000, an M.Sc. in medical radiation physics from the University of Wales Swansea in the UK, and a Ph.D. in health sciences from Swansea University in the UK in 2012. His background in diagnostic radiology as a physicist has helped him to deliver courses related to radiation, such as radiation physics, medical imaging, medical radiation, and physics applications in medicine, radiation protection, CT, nuclear medicine, and radiotherapy physics. His research interests are in radiation, medical radiation, radiation protection, medical physics, nuclear medicine, and health sciences. He can be contacted at matari@bau.edu.jo.

Ghaith M. Jaradat received a B.Sc. degree in computer science from Jerash University, Jordan, in 2004, and his M.Sc. degree in intelligent systems from Utara University, Malaysia, in 2007, and a Ph.D. degree in computer science from the National University of Malaysia, Malaysia, in 2012. He is currently an associate professor at Amman Arab University, Jordan since 2020. His research interests are mainly directed to metaheuristics and combinatorial optimization problems including timetabling, routing, quadratic, and rostering. His research interests include the applications of artificial intelligence, including deep learning, evolutionary and heuristic optimization techniques to power system planning, operation and control, text classification, and feature selection prediction models. He can be contacted at g.jaradat@aau.edu.jo.

Hemn Barzan Abdalla holds a Ph.D. degree in the field of communication and information engineering. He possesses one decade of experience in teaching and worked as a project assistant in various higher education places. In addition, currently, he is working as an assistant professor at Wenzhou-Kean University as a member of the Institute of Training and Development in Sulaimani (KRG). He is an editorial board member/reviewer of international/national journals and conferences. He has 13+ years of experience in the IT industry and he has more than 100 project systems for several places. His research interests include big data, data security, NoSQL, database management system, software engineering, and application and software testing. He can be contacted at habdalla@kean.edu.

Abdallah Mohd Abualkishik is currently an assistant professor at the Faculty of Computing and Information Technology, Suhar University. He received his B.S. degree in computer science in 2004 from Jarash University, Jordan, his M.Sc. degree in intelligent systems in 2007 from North Malaysia University (UTARA), Malaysia, and his Ph.D. in artificial intelligence and software engineering in 2011 from Malaysia National University (UKM), Malaysia. He has published dozens of research papers in different fields of computer science. His research interests include artificial intelligence, information technologies, internet of things, and smart systems to solve real-life and community problems. He can be contacted at aabualkishik@su.edu.om.

Mutasem Khalil Alsmadi is currently an associate professor at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University. He received his B.S. degree in software engineering in 2006 from Philadelphia University, Jordan, his M.Sc. degree in intelligent systems in 2007 from the University Utara Malaysia, Malaysia, and his Ph.D. in computer science from The National University of Malaysia. He has published more than one hundred papers in the image processing and algorithm optimization areas. His research interests include artificial intelligence, pattern recognition, algorithm optimization, and computer vision. He can be contacted at mksalsmadi@gmail.com.