CrossMark

REGULAR PAPER

Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines

Syed Muhammad Saqlain[1] · Muhammad Sher[1] · Faiz Ali Shah[2] ·
Imran Khan[1] · Muhammad Usman Ashraf[1,3] · Muhammad Awais[4] ·
Anwar Ghani[1]

Abstract Heart is one of the essential operating organs of the human body and its failure is a major contributing factor toward the human deaths. Coronary heart disease may be asymptotic but can be anticipated through the medical tests and daily life routine of the subject. Diagnosis of the coronary heart disease needs a specialized medical resource with the plenty of experience. All over the world and particularly in the developing countries, there is a lack of such experts which make the diagnosis more difficult. In this paper, we present a clinical heart disease diagnostic system by proposing feature subset selection methodology with an object of achieving improved performance. The proposed methodology presents three algorithms for selecting candidate feature subsets: (1) mean Fisher score-based feature selection algorithm, (2) forward feature selection algorithm and (3) reverse feature selection algorithm. Feature subset selection algorithm is presented to select the most decisive subset from the candidate feature subsets. The features are added to the feature subsets on the basis of their individual Fisher scores, while the selection of a feature subset depends on its Matthews correlation coefficient score and dimension. The selected feature subset with the reduced dimension is fed to the RBF kernel-based SVM which results in binary classification: (1) heart disease patient and (2) normal control subject. The proposed methodology is validated through accuracy, specificity and sensitivity using four UCI datasets, i.e., Cleveland, Switzerland, Hungarian and SPECTF. The statistical results achieved using the proposed technique are shown in comparison with the existing techniques reflecting its better performance. It has an accuracy of 81.19, 84.52, 92.68 and 82.7% for Cleveland, Hungarian, Switzerland and SPECTF, respectively.

✉ Syed Muhammad Saqlain
  syed.saqlain@iiu.edu.pk

1 Department of Computer Science & Software Engineering, International Islamic University Islamabad, Islamabad, Pakistan

2 University of Tartu, Tartu, Estonia

3 IBMS, University of Agriculture Faisalabad, Faisalabad, Pakistan

4 Government College University Faisalabad, Faisalabad, Pakistan

Springer

1 Introduction

In every human body, there are blood vessels that span an approximate length of 60,000 miles while ensuring the proper functionalities of all its organs. The heart regularly pumps fresh blood into these blood vessels, at a rate of 4.7 l per minute. As such, the heart may be called as the engine of the human body whose failure results in the collapse of the body. As per reports by the World Health Organization, cardiac diseases are one of the main factors contributing toward human deaths. Different surveys held since 2012 concluded among almost 56 million people lost their lives in 2012; 17.5 million were caused by cardiovascular diseases [1]. It was the highest single natural factor responsible for the loss of lives. Heart diseases can be cured or controlled if they are diagnosed in their early stages. This is possible through a series of medical tests conducted by specialized medical human experts and coupled with the medical history of the subject. As per World Bank figures, the ratio of general physicians to living humans in developed countries is 3:1000, while in the developing countries this ratio is 0.6:1000 [2]. As there are fewer heart specialists than general physicians, we can say that the availability of the former is far lower than the latter. Particularly for the subjects living in remote rural areas it nearly becomes impossible in approaching a specialist.

Automated computer systems have amazingly improved the quality and standard of the life in every aspect. These systems are also facilitating medical professionals for accomplishing their jobs. Along with the development of automatic diagnosis systems for other diseases, research is in progress to diagnose heart diseases through computerized intelligent systems [3]. During the last two decades, considerable work is presented for achieving the desired results. In these systems, statistics of medical reports, medical history and the subject's life habits are provided as inputs. The automated system, through its learning and decision functions, then predicts whether the subject is suffering from heart disease or not. Research on these systems have focused on two areas: (1) the selection of important features while discarding redundant ones and (2) the selection of an appropriate classifier. The selection of the appropriate features subset not only reduces the computational complexity but improves the classification results [4]. Features should be selected in a way in which each can be considered as a discriminating feature. The combination of all the individual discriminating features must then produce a better classification. The role of the classifiers is the key to the success of prediction systems. In the broader aspects, two types of machine learning techniques are used, i.e., supervised and unsupervised learning-based techniques. Supervised learning is quite successful while the historical data along with the class labels are available. Support vector machines (SVMs) are supervised learning techniques which are used to solve both linear and nonlinear problems. When radial basis function (RBF)-based kernels are used in SVM, it makes them capable of solving the linearly non-separable problems.

This paper presents feature selection techniques which select the features on the basis of their individual and collective performance of the subset when added to it. It not only considers the features with individual high discriminating scores but the ones with lower scores are taken into account as well if their presence improves the classification accuracy for the whole subset. The Matthews correlation coefficient (MCC) is used as a metric for selecting the feature subset. The Matthews correlation coefficient (MCC) is used as a metric for selecting the feature subset. The higher the MCC score, the better the feature subset.

RBF kernel-based SVM classifier takes the reduced dimensional feature subset as an input and predicts whether the subject under consideration is a potential heart disease patient or a normal control subject. The rest of the paper is organized as follows: Sect. 2 briefly discusses the latest relevant literature; Sect. 3 presents the proposed research; Sect. 4 discusses the datasets and evaluation criteria used; Sect. 5 presents, discusses and analyzes the research results; and Sect. 6 concludes the paper.

2 Literature review

This section presents a review of the relevant research contributions published in recent past. Long et al. [5] proposed a heart disease diagnostic system which used fuzzy-based classification. The authors used a feature subset by reducing the original dimension of the feature set through rough sets and the chaos firefly algorithm. The technique was validated by testing it over the heart disease and SPECF datasets. Extreme learning-based algorithm was used by Ismaeel et al. [6] for heart disease diagnostics. Their research had achieved 80% accuracy by implementing their methodology over the Cleveland dataset. Minimum distance-based KNN classifier was used by Krishnaiah et al. [7] to diagnose the heart diseases. As an initial step, they fuzzified medically measured data and then provided it to the KNN through fuzzy membership. An intelligent prediction system was proposed by Chitra and Seenivasagam [8] which used the feed forward neural network and the cascaded correlation neural network for classification. Srinivas et al. [9] generated fuzzy rules through set theory which were then used to perform fuzzy classification. The technique was validated through results which were obtained by implementing it over the UCI dataset (i.e., Cleveland, Hungarian, SPECTF and Switzerland). The authors achieved the highest accuracy (80%) for the Switzerland dataset, while they achieved the lowest accuracy (42%) for the Hungarian dataset. A hybrid approach was developed by Yang et al. [10] which combined adaptive network-based fuzzy inference system (ANFIS) with linear discriminant analysis (LDA) and was used to assess the risk of coronary heart disease. The technique was validated by experimenting it over a dataset of the Korean National Health and Nutrition Examinations Survey, achieving an accuracy of 80.2%. A hybrid modeling scheme for the heart disease diagnosis was proposed by Yuehjen et al. [11]. It was a two-stage methodology in which initially a set of explanatory variables were reduced through the logistic regression (LR), multivariate adaptive regression splines (MARS) and rough set (RS) followed by a classification through the artificial neural network (ANN). An expert system for the diagnosis of coronary heart disease was presented by Muthukaruppan and Er [12] in which they employed particle swarm optimization (PSO) along with fuzzy classifiers. Fuzzy membership functions were tuned using the PSO. Afterward, they used the Cleveland and Hungarian heart disease datasets for validating their technique.

Anooj [13,14] presented a clinical decision support system (CDSS) for heart disease diagnosis. The proposed CDSS had two stages: In the first stage, attribute selection and attribute weightage were achieved using the data mining techniques while Fuzzy rule-based CDSS was proposed in the second stage. Comparative results were presented along with the ANN-based technique over UCI datasets. Shilaskar and Ghatol [15] used a reduced feature subset for the diagnosis of heart disease. They used distance measures as a metric to select features through forward inclusion, forward selection and backward elimination search techniques. To improve the classification accuracy, the authors proposed their own hybrid forward selection algorithm and achieved comparably better results than the forward selection and backward elimination algorithm. Expert judgment was compared with automated feature selection

techniques using different classifiers [16]. The authors showed that a combination of expert judgment with intelligent feature selection techniques improved the accuracy of the naive Bayes, IBK and SMO classifiers. ECG heart rate signals were used as input to predict cardiac health by Giri et al. [17]. Frequency sub-bands of the input signals were obtained using discrete wavelet transform (DWT). Later on, principal component analysis (PCA), independent component analysis (ICA) and linear discriminant analysis (LDA) were employed over DWT coefficients. Different tests using the Gaussian mixture model (GMM), support vector machine (SVM), probabilistic neural network (PNN) and artificial neural network (ANN) classifiers were employed. Zhao et al. [18] presented a feature selection framework which was used to preserve similarity measures through a conventional combinatorial optimization formulation. In order to improve the efficiency and effectiveness of their methodology, they extended it with a sparse multiple-output regression formulation. The authors concluded that it allowed them to achieve the better results.

The feature selection problem was investigated by Hancer et al. [19] through proposing ant–bee colony (ABC)-based algorithm. They extended the standard ABC algorithm by integrating it with evolutionary-based similarity mechanism. A feature selection criteria was proposed by Zhang et al. [20] where they incorporated the $l_{2,1}$ norm regularization into the original Fisher criterion. It was assured that the $l_{2,1}$ norm regularization term would achieve the sparsity of the feature selection matrix resulting in the feature selection as a globally optimized solution. A multiclass problem was solved by Li et al. [21] who proposed a cost-sensitive LDA classifier. Linear discriminant coefficients were used for estimating the posterior probabilities of state-of-the-art testing instances. The authors achieved efficiency in terms of low computational cost and higher accuracy than the existing state-of-the-art methodologies. Markos et al. [22] proposed a four-staged fuzzy-based classifier. A Generalized Fisher score was proposed by Gu et al. [23] which maximized the lower bound of the traditional Fisher score for selecting the feature subset. The resulting selected features were reformulated as a quadratically constrained linear programming (QCLP) and solved it by a cutting plane algorithm. Information gain and divergence were used by Lee and Lee [24] for selecting feature subsets. The authors reduced the redundancy among features and maintained information gain through this methodology.

Support vector machines (SVMs) are the collection of supervised learning methodologies [25]. Akay [26] presented their technique that was used to predict women breast cancer disease. The authors utilized the classification ability of SVM using reduced dimensional feature set. They achieved 99.51% accurate results. Çomak et al. [27] used Doppler signals of the heart valves for predicting the heart disease. They used least-squares support vector machine (LS-SVM) as a predictor. Yahiaoui et al. [28] used SVM for diagnosis of tuberculosis disease. Comparative results of the proposed technique with the existing techniques were presented, and the proposed methodology reported better results. Huang et al. [29] used ensembles of SVM classifiers for predicting breast cancer disease. The authors discussed that ensembles of SVM produced better results than the single SVM, and while using different kernels, RBF kernels performed better than all its fellows.

3 Proposed methodology

The proposed technique selects the feature subset by proposing feature selection algorithms. The selected subset then is used to train and test RBF-based SVM classifier (Fig. 1).

Prior to the feature selection process, feature attribute ranges are standardized and Fisher score of each feature is calculated. During the feature selection procedure, we present mean

Fisher score-based feature selection algorithm (MFSFSA), forward feature selection algorithm (FFSA) and reverse feature selection algorithm (RFSA) for subset selection. The feature subset selection algorithm (FSSA) returns the most effective lower-dimensional feature subset.

Definition 1 Heart feature set is S_{HP}, set of suspected heart patients, having M Tuples. Each tuple T_i has N features. S_{HP} is represented by $S_{HP} = \{T_1, T_2, \ldots, T_M\}$. Tuple T_i is represented by $T_i = \{f_1, f_2, \ldots, f_N\}$ for every $1 \leq i \leq M$.

We need to classify each $T_i \in C$ where $C = \{C_{HDP}, C_{NCS}\}$ and C_{HDP} and C_{NCS} represent class of heart disease patient (HDP) and class of normal control subject (NCS), respectively.

3.1 Feature selection

To avoid the participation of redundant and the lesser class discriminating features, whose presence not only compromises the classification accuracy but leads toward high computational complexity as well, a feature subset with higher class discriminating capabilities must be selected. The required goal is achieved by performing the following steps:

3.1.1 Data standardization

The medical data of a subject consist of multiple attributes, represented by the different data types. Some of the attributes are of binary nature, while others may represent decimal or fractional numbers. The ranges of all the attributes can be quite diverse resulting in a bias toward selection of the certain features.

Definition 2 A Tuple T with features f_i, $1 \leq i \leq n$, each f_i may have different data format. Data standardization is the process of transforming diverse data representations into a unique format. The transformed single format helps in comparing and classifying the data instances.

As a means of preprocessing, all the features are standardized using following relation:

$$f_i' = \frac{f_i - \overline{f_i}}{\sigma_{f_i}} \tag{1}$$

where f_i is the ith feature in its raw form, $\overline{f_i}$ is the mean of the ith feature, while σ_{f_i} is the standard deviation of the ith feature. f_i' is the ith feature in a standardized form.

3.1.2 Fisher score calculation

The feature selection methodology should favor the features with higher discriminating capabilities while the others should be suppressed.

Definition 3 Using a feature set $F = \{x_1, x_2, \ldots, x_N\}$, Fisher score is used to measure the class discriminant nature of each x_i. Features with the higher Fisher scores are more discriminant than the features with lower ones.

The proposed technique uses Fisher score to rank each feature which is calculated by using following mathematical formulation:

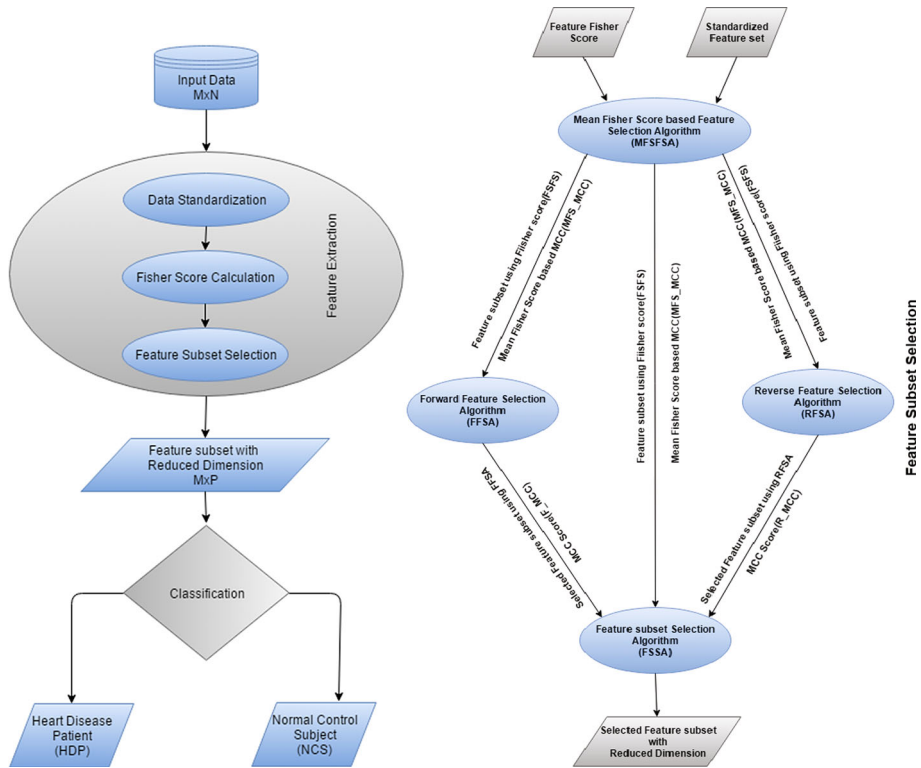$$SCF(F_i) = \frac{\sum_{j=1}^{C} \eta_j (\mu_{i,j} - \mu_i)^2}{\sum_{j=1}^{C} \eta_j \sigma_{i,j}^2} \tag{2}$$

Fig. 1 Flow diagram of proposed technique

where η_j, $\mu_{i,j}$, μ_i and $\sigma_{i,j}^2$ represent the number of tuples lying in the jth class, mean of the ith feature in the jth class, mean of the ith feature and variance of the ith feature in the jth class, respectively. C represents total number of classes.

3.1.3 Feature subset selection

In order to select a minimal number of features used for coronary heart disease diagnosis, we present a set of feature selection algorithms. The final outcome of the proposed set of algorithms is a P-dimensional feature subset, i.e., $M \times N \rightarrow M \times P$, where M, N and P represent the number of instances in the dataset, original feature dimension and the reduced feature dimension, respectively.

3.2 Mean Fisher score-based feature selection algorithm (MFSFSA)

Definition 4 Given a set of ranked Fisher score, $RFS = \{fs_1, fs_2, \ldots, fs_m\} \forall fs_{j-1} \geq fs_j$, and feature set corresponding to RFS, $FSR = \{x_1, x_2, \ldots, x_m\} \forall fs_{x_{j-1}} \geq fs_{x_j}$, mean Fisher score-based feature selection algorithm (MFSFSA) selects a feature subset based on individual Fisher scores in comparison with the mean Fisher score. It returns the MCC score of the selected feature subset as well.

Algorithm 1 : Mean Fisher Score-based Feature Selection Algorithm (MFSFSA)

INPUT: Ranked Fisher Score (RFS), Feature Set corresponding to Ranked Fisher Score (FSR)
OUTPUT: Feature subset using FS (FSFS), Mean Fisher score-based Matthew's Correlation Coefficient (MFS_MCC)

1: Initialize FSFS = { φ }
2: Initialize Matthew's Correlation Coefficient (MCC), MFS_MCC=0
3: Calculate mean Fisher score (μ_{fs}) using $\mu_{fs} = \frac{\Sigma_{i=1}^{N} fs_i}{N}$
4: Add all features,x_j from FSR to FSFS $\forall fs_j \geq \mu_{fs}$
5: Train classifier using FSFS
6: Through Validation obtain MCC using: MFS_MCC $= \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FN)(TN+FP)}}$, where
   TP, *TN*, *FP* and *FN* represent true positive, true negative, false positive and false negative, respectively.

In MFSFSA, all the features having Fisher score higher than the mean score are selected in feature subset as they have more class discriminative powers. The selected feature subset is used for training support vector machines and calculating the MFS_MCC through a validation process.

3.2.1 Forward feature selection algorithm (FFSA)

Definition 5 Given a set of ranked Fisher score, i.e., $RFS = \{fs_1, fs_2, \ldots, fs_N\} \forall fs_{j-1} \geq fs_j$, feature set corresponding to RFS, i.e., $FSR = \{x_1, x_2, \ldots, x_N\} \forall fs_{x_{j-1}} \geq fs_{x_j}$, feature subset using MFSFSA (FSFS), i.e., $FSFS = \{x_1, x_2, \ldots, x_Q\} Q \leq N$ and MCC score through MFSFSA (MFS_MCC), forward feature selection algorithm (FFSA) selects and adds features in $FSFS$ as per their Fisher score (descending order) and makes it a potential SFS_{F_i}, i.e., feature subset. It returns the selected feature subset with the maximum MCC score.

Algorithm 2 : Forward Feature Selection Algorithm (FFSA)

INPUT: Ranked Fisher Score (RFS), Feature Set corresponding to Ranked Fisher Score (FSR),
Feature subset using MFSFSA (FSFS), MCC score through MFSFSA(MFS_MCC)
OUTPUT: Selected feature subset through Forward Feature Selection Algorithm (SFS_F),
MCC score corresponding to selected feature subset (F_MCC)

1: $\delta_f = $ MFS_MCC$\times 0.15$
2: TS $= MFS\_MCC$
3: Initialize $SFS_F = FSFS$
4: **repeat**
5:     $SFS_F = SFS_F \cup \{x_i\}$.
6:     Train classifier using updated SFS_F
7:     validate train data and obtain Matthew's Correlation Coefficient, i.e., F_MCC
8:     **if** $F\_MCC > TS + \delta_f$ **then**
9:        $TS = F\_MCC$
10:        Keep x_i in SFS_F
11:        $\delta_f = \delta_f / 2$
12:    **end if**
13: **until** $fs_i < \mu_{fs} \ \forall x_i \in$ FSR
14: return SFS_F, F_MCC

During processing of FFSA, each feature having a score lesser than the mean Fisher score is added to feature subset. The process is accomplished by adding features in descending order with respect to their corresponding Fisher scores. The classifier is trained and validated using an updated feature subset. A corresponding MCC is then calculated. If the selection of a feature increases the Matthews correlation coefficient score by an acceptable amount, the feature is kept in a list of selections; otherwise, it is discarded. The process continues until all the features are evaluated.

3.2.2 Reverse feature selection algorithm (RFSA)

Definition 6 Given a set of ranked Fisher score, i.e., $RFS = \{fs_1, fs_2, \ldots, fs_N\} \forall fs_{j-1} \geq fs_j$, feature set corresponding to RFS, i.e., $FSR = \{x_1, x_2, \ldots, x_N\} \forall fs_{x_{j-1}} \geq fs_{x_j}$, feature subset using MFSFSA (FSFS), i.e., $FSFS = \{x_1, x_2, \ldots, x_Q\} Q \leq N$ and MCC score through MFSFSA (MFS_MCC), reverse feature selection algorithm (RFSA) selects and adds features in $FSFS$ as per their Fisher score (ascending order) and makes it a potential SFS_{R_i}, i.e., feature subset. It returns the selected feature subset with the maximum MCC score.

Algorithm 3 : Reverse Feature Selection Algorithm (RFSA)

INPUT: Ranked Fisher Score (RFS), Feature Set corresponding to Ranked Fisher Score (FSR),
Feature subset using MFSFSA (FSFS), MCC score through MFSFSA(MFS_MCC)
OUTPUT: Selected feature subset through Reverse Feature Selection Algorithm (SFS_R),
MCC score corresponding to selected feature subset (R_MCC)

1: TS $= MFS\_MCC$
2: $\delta_r = $ TS$\times 0.15$
3: Initialize $SFS_R = FSFS$
4: **repeat**
5:     $SFS_R = SFS_R \cup \{y_j\}$
6:     Train classifier using updated SFS_R
7:     validate train data and obtain Matthew's Correlation Coefficient, R_MCC.
8:     **if** R_MCC $> TS + \delta_r$ **then**
9:         $TS = R\_MCC$
10:        $\delta_r = \delta_r/2$
11:    **end if**
12: **until** $\forall y_j \in$ FSR: Starting from smallest Fisher score to highest
13: return SFS_R, R_MCC

The FFSA favors the features with higher Fisher scores by giving them chance prior to the features with lower scores. To avoid the biases, RFSA works in reverse to the FFSA by chancing the features with lower Fisher score first.

3.2.3 Feature subset selection algorithm (FSSA)

Definition 7 Given the MCC scores from MFSFSA, FFSA and RFSA along with their corresponding feature subsets, feature subset selection algorithm (FSSA) computes and returns the selected feature subset (SFS).

Algorithm 4 : Feature subset Selection Algorithm (FSSA)

INPUT: FSFS, SFS_F, SFS_R
MFS_MCC, F_MCC, R_MCC
OUTPUT: selected feature subset (SFS)

1: Get the maximum of $\{F\_MCC, R\_MCC, MFS\_MCC\}$ representing Highest Matthew's Correlation Coefficient.
2: **if** All are equal OR MFS_MCC is the maximum **then**
3:    SFS = FSFS
4:    return SFS
5: **end if**
6: **if** F_MCC is the maximum **then**
7:    SFS = SFS_F
8:    return SFS
9: **end if**
10: **if** R_MCC is maximum **then**
11:    SFS = SFS_R
12:    return SFS
13: **end if**
14: **if** $R\_MCC \equiv F\_MCC$ and $R\_MCC, F\_MCC > MFS\_MCC$ **then**
15:    check SFS_F, SFS_R having lesser number of features in them
16:    **if** SFS_R have lesser number of features **then**
17:      SFS = SFS_R
18:      return SFS
19:    **end if**
20:    **if** SFS_F have lesser number of features **then**
21:      SFS = SFS_F
22:      return SFS
23:    **end if**
24:    **if** SFS_F and SFS_R have the same length **then**
25:      SFS = SFS_F
26:      return SFS
27:    **end if**
28: **end if**

At the end, FSSA looks for the feature subset that results in highest Matthews correlation coefficient score. If FFSA and RFSA do not improve the MCC score, then subset obtained through MFSFSA gets selected. Otherwise, feature subset scoring the highest MCC is selected. If MCC scores for both the subsets (FFSA and RFSA) are equal, then the feature set obtained through the FFSA is selected as its features are more probable of having the higher Fisher scores.

Algorithm 5 : Cardiac System Algorithm (CSA)

1: (FSFS, MFS_MCC)=MFSFSA(RFS,FSR)
2: **if** $MFS\_MCC \equiv 1$ **then**
3:    return FSFS
4: **else**
5:    (SFS_F , F_MCC)=FFSA(RFS,FSR,FSFS,MCC)
6:    (SFS_R , R_MCC)=RFSA(RFS,FSR,FSFS,MCC)
7:    SFS=FSSA(FSFS, MFS_MCC , SFS_F , F_MCC , SFS_R , R_MCC)
8: **end if**

The cardiac system algorithm (CSA) presents the flow of all the four proposed algorithms.

Theorem 1 *The FSSA returns the minimum item feature subset having maximum class discriminating ability.*

Proof Let F be a set of m features, i.e., $F = \{x_1, x_2, x_3, \ldots, x_m\}$, F' is the set having Fisher scores for each $x_j: 1 \leq j \leq m$. Algorithm 1, MFSFSA, returns a feature subset with features having individual Fisher scores greater than mean Fisher score, i.e., $FSFS = \{x_i, x_{i+1}, \ldots x_s\}$, $1 \leq s \leq m$ and MFS_MCC as its class discriminant score.

1. If MFS_MCC is one, FSFS is the most efficient feature subset and no further computation is required.
2. If MFS_MCC is less than one, then FSFS may not be an optimized subset. Algorithm 2 returns SFS_F, a selected feature subset with the higher class discriminant score, i.e., $F\_MCC \geq MFS\_MCC$ and $FSFS \subset SFS\_F$. From Algorithm 3, it may be observed that SFS_R, a feature subset may be obtained having higher class discrimination, i.e., $R\_MCC \geq MFS\_MCC$ and $FSFS \subset SFS\_R$.
3. By using Algorithm 4, if MFS_MCC, F_MCC and R_MCC are equal then feature subset corresponding to MFS_MCC is selected as it has the lower dimension while sharing higher class discriminant score among others. Otherwise, the feature subset with higher discriminant score will be selected.
4. Through step 3, if F_MCC and R_MCC are equal, resulting in both the corresponding subsets with the equal class discriminant scores, a subset with the lower dimension would be selected. It gives a feature subset with the higher class discriminant score and lower possible feature dimension.
5. If step 3 returns equal F_MCC and R_MCC scores and dimensions of their corresponding feature subsets, i.e., SFS_F and SFS_R, are equal as well, then SFS_F should be selected as it has a same class discriminant score as SFS_R, but its individual features are more probable of having a higher class discriminant score.

From 1, 3, 4 and 5, it is clear that the FSSA always returns a feature subset with the highest class discriminating ability and the minimum possible dimension. □

3.3 Classification

In the proposed technique, support vector machine (SVM) is used for binary classification (i.e., heart disease patient and normal control subject). The input dataset $M \times P$ is divided into two subsets: (1) train data and (2) test data. Train data are of order $T_r \times P$, and test data have an order $T \times P$. Decision functions for linear SVMs can be represented by $(u.s_j + c) \leq -1$ if $t_j = -1$ and $(u.s_j + c) \geq 1$ if $t_j = 1$. The two relations can be combined to cover both cases, i.e., $t_j(u.s_j + c) \geq 1$ where t represents the class, s is the input data, u represents weight and c is the margin. The problem of coronary heart disease anticipation comes under the class of nonlinear problems. Decision function for the nonlinear classification is represented as $f(s) = u.\Phi(s) + c$. This is presented as we are dealing with the research problem having input feature subset instead of a single-dimensional problem. For a higher-dimensional data convergence, SVM kernel functions are widely used. In a kernel-based SVM, u may be represented as $u = \sum_{j=1}^{m} \beta_j \Phi(s_j)$. Decision function for a nonlinear classification problem having multidimensional feature set can be represented as:

$$f(s) = \sum_{j=1}^{m} \beta_j \Phi(s_j).\Phi(s) + c \tag{3}$$

where $\Phi(s_j).\Phi(s)$ is defined as the kernel function and denoted as $KF(s, s_j)$. By replacing $\Phi(s_j).\Phi(s)$ by $KF(s, s_j)$, the decision function is given as:

$$f(s) = \sum_{j=1}^{m} \beta_j \cdot KF(s, s_j) + c \tag{4}$$

In order to solve the nonlinear separable problems (with high-dimensional feature set), the radial basis function-based kernel is identified as the most suitable. This is due to the RBF kernel being known to be used in solving infinite-dimensional problems. The RBF kernel can be represented as

$$KF(s, s_j) = e^{(-\gamma \|s - s_j\|^2)} \tag{5}$$

By substituting Eq. (5) into Eq. (4), the required decision function becomes:

$$f(s) = \sum_{j=1}^{m} \beta_j \cdot e^{(-\gamma \|s - s_j\|^2)} + c \tag{6}$$

Theorem 2 *The RBF-based SVM solves the heart disease classification problem.*

Proof Given an input data with $M \times P$ dimension partitioned into two subsets, training data ($S \times P$) and test data ($T \times P$). Training and testing a non-linearly separable feature set (multidimensional) having binary classes are a nonlinear problem. In order to produce a nonlinear boundary projection, SVM uses $f(s) = u.\Phi(s) + c$ as decision function. The diagnosis of the heart disease through feature subset is a problem of multidimensional spaced data. As multidimensional feature set classification is achieved using kernel-based decision function, so using $f(s) = \sum_{j=1}^{m} \beta_j \Phi(s_j).\Phi(s) + c$, kernel-based decision function solves the nonlinear classification problem. RBF kernel, $KF(s, s_j) = e^{(-\gamma \|s - s_j\|^2)}$, is proved in solving infinite-dimensional problems. So by using $f(s) = \sum_{j=1}^{m} \beta_j.e^{(-\gamma \|s - s_j\|^2)} + c$, classification of nonlinear separable multidimensional feature set can be achieved. Hence, it is concluded that the RBF-based SVM solves the heart disease classification problem. □

4 Experimental setup

4.1 Evaluation criteria

In order to evaluate the performance of the proposed technique, three qualitative measurements have been used, i.e., (1) accuracy, (2) specificity and (3) sensitivity.

$$\text{Sensitivity} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \tag{7}$$

$$\text{Specificity} = \frac{\text{True Negative}}{(\text{True Negative} + \text{False Positive})} \tag{8}$$

$$\text{Accuracy} = \frac{(\text{True Negative} + \text{True Positive})}{(\text{True Negative} + \text{True Positive} + \text{False Negative} + \text{False Positive})} \tag{9}$$

where

True positive: heart disease patient correctly identified as heart disease patient
False positive: normal control subject incorrectly identified as heart disease patient

True negative: normal control subject correctly identified as normal control subject
False negative: heart disease patient incorrectly identified as normal control subject

4.2 Dataset

Four UCI [30] datasets are used: (1) Cleveland, (2) Hungarian, (3) Switzerland and (4) SPECTF. Originally, the first three datasets had 76 features out of which 14 attributes including class label are selected, while SPECTF have 45 attributes which are used in total. Details of the used attributes(UA) for evaluating the proposed technique are shown in Table 1.

Three distinct statistics have been shown for each of the dataset (i.e., the total number of records, training and test data statistics). The Cleveland dataset with 303 instances is partitioned into training and test dataset with 202 and 111 instances, respectively. The said process is achieved with an intuition of keeping symmetry in class membership statistics. Out of the 202 training records, 111 have membership in the HDP class while 91 instances belong to NCS. From the test data subset, 53 records are part of HDP class and 48 records belong to NCS. The rest of the datasets are analyzed in the same way as well.

5 Results and discussion

The detail of the dataset [30] used for experimentation is explained in Tables 1 and 2. In order to avoid the bias in a comparative analysis with the existing techniques, the dataset partitioning is performed in the same way as opted by the competitive techniques.

Table 3 shows the mean Fisher score, dimension of the intermediate selected subsets using MFSFSA, FFSA, RFSA and finally selected subset using FSSA. Originally three of the datasets have 13-dimensional feature set (i.e., Cleveland, Hungarian and Switzerland) while SPECTF has 44 features. The mean Fisher score varies from 0.0758 (Switzerland) to 0.1457 (Cleveland). For the Cleveland dataset, six features are selected using MFSFSA and one extra feature is selected using the FFSA and RFSA. The FSSA selects subset produced by the FFSA. While experiments are performed over the Hungarian dataset, MFSFSA selects four features while two and three extra features are selected through the FFSA and RFSA, respectively. In the same manner, four features are selected by the MFSFSA for the Switzerland dataset. Afterward, the FFSA selects four, while the RFSA selects two extra features. The FSSA selects the feature subset produced by the FFSA. Experimental results of the proposed technique over the SPECTF showed the selection of a 19-dimensional feature subset. All the 19 features are selected using the MFSFSA.

Table 4 clearly shows a reduction in the original feature dimension using the proposed technique. Dimensions of the selected feature subsets for Cleveland, Hungarian, Switzerland and SPECTF are 47, 47, 39 and 57%, respectively, which are less than their original feature sets.

Figure 2 shows a graphical representation of the performance achieved by the proposed technique. It shows the performance of the proposed technique in terms of accuracy, sensitivity and specificity.

5.1 Results for the Cleveland dataset

Figure 3a, b, c graphically shows the feature selection statistics using FFSA, RFSA and FSSA for the Cleveland dataset. It shows that out of the 13 features of Cleveland dataset, MFSFSA selects six features with the Fisher score greater than the mean Fisher score of all the features.

Table 1 Partition details of the datasets in train and test sets

Dataset	No. of records	Training data			Test data		
		Total training records	Heart patient records	Non-heart patient records	Total training records	Heart patient records	Non-heart patient records
Cleveland	303	202	111	91	101	53	48
Hungarian	283	142	89	53	141	95	46
Switzerland	123	82	6	76	41	2	39
SPECTF	267	80	40	40	187	15	172

Table 2 Dataset specifications

UA specifications for Cleveland, Hungarian and Switzerland

Att. No.	Attribute	Description	Data representation	Attribute specification
1	Age	Age in years	Numerical	
2	Sex	Sex	Binary	1 = male
				0 = female
3	CP	Chest pain	Numeric (1–4)	1 = typical angina
				2 = atypical angina
				3 = non-anginal pain
				4 = asymptomatic
4	Trestbps	Resting blood pressure in mm Hg	Numeric	
5	Chol	Serum cholestoral in mg/dl	Numeric	
6	Fbs	Fasting blood sugar > 120 mg/dl	Binary	1 = true
				0 = false
7	Resteg	Resting electrocardiographic results	Numeric (0–2)	0=normal
				1=having ST-T wave abnormality
				2=left ventricular hypertrophy
8	Thalach	Maximum heart rate achieved	Numeric	
9	Exang	Exercise induced angina	Binary	1 = yes
				0 = No

Table 2 continued

UA specifications for Cleveland, Hungarian and Switzerland

Att. No.	Attribute	Description	Data representation	Attribute specification
10	Oldpeak	ST depression induced by exercise relative to rest	Floating Numeric	
11	Slope	The slope of the peak exercise ST segment	Numeric (1–3)	1=upsloping
				2=flat
				3=downsloping
12	Ca	Number of major vessels	Numeric (0–3)	
13	Thal	Type of defect	Numeric	3=normal
14	Num	Diagnosis of heart disease	Binary	Value 0: < 50% diameter narrowing(Normal)
				Value 1: > 50% diameter narrowing(Patient)

UA Specifications for SPECTF

Att. No.	Attribute	Data representation	Data values
1	OVERALL_DIAGNOSIS	Binary	0:Normal1:Patient
2–45	F1R,F1S,F2R,F2S,F3R,F3S, F4R,F4S,F5R,F5S,F6R,F6S,F7R, F7S,F8R,F8S,F9R,F9S,F10R,F10S, F11R,F11S,F12R,F12S,F13R,F13S, F14R,F14S,F15R,F15S,F16R, F16S,F17R,F17S,F18R,F18S, F19R,F19S,F20R,F20S,F21R, F21S,F22R,F22S	Numeric	0–100

Table 3 Results of proposed technique over all datasets

Dataset	Original dimension	Mean score	Fisher Feature subset dimension through MFSFSA	Feature subset dimension through FFSA	Feature subset dimension through RFSA	Feature subset through FSSA	Classification accuracy (%)
Cleveland	13	0.1457	6	7	7	FFSA	81.19
Hungarian	13	0.1193	4	6	7	RFSA	84.52
Switzerland	13	0.0191	4	8	6	FFSA	92.68
SPECTF	44	0.0758	19	NA	NA	MFSFSA	82.70

Table 4 Feature dimension details

Dataset	Original dimension	Dimension of selected feature subset	Dimension reduced (%)
Cleveland	13	7	47
Hungarian	13	5	47
Switzerland	13	8	39
SPECTF	44	19	57
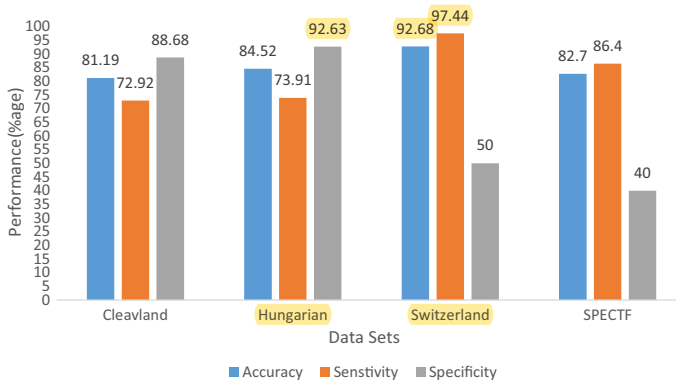
Fig. 2 Performance analysis of the proposed technique

The MCC score (i.e., MFS_MCC) for the selected feature subset is 0.79. Figure 3a shows the MCC score for the feature subset selection using FFSA (i.e., F_MCC). Initially, δ_f is the 15th percent of the MFS_MCC score. The inclusion of the 7th ranked feature increased the MCC score from 0.79 to 0.90 and fulfills the required criteria, i.e., $F\_MCC > TS + \delta_f$. After adding the 8th ranked feature to the selected subset, the F_MCC score achieved is 0.93, an increase of 0.03. The 8th ranked feature cannot be selected as the F_MCC score should have been boosted by 0.06. All the other features are added and tested against MCC boosting criteria, but none of the features is selected in feature subset. Through the FFSA, only the 7th ranked feature is selected. This becomes a member of existing feature subset obtained through the MFSFSA criteria.

Figure 3b shows the MCC scores of selected feature subsets through the RFSA. As with the case of the FFSA, six features are selected through the MFSFSA. Rest of the features are added to the selected subset in reverse to the FFSA (i.e., features with low Fisher score are tested first). When the 13th ranked feature (i.e., a feature with the lowest Fisher score) is added to the feature subset, R_MCC score is boosted from 0.79 to 0.85. This corresponds to an increase of 0.06 points. The minimum threshold increase in the score should be 0.11 for the inclusion of the first feature fulfilling $R\_MCC > TS + \delta r$, where δ_r is initialized with $0.15 \times MFS\_MCC$. This feature cannot be selected. The process continues and when the 8th ranked feature is added to the feature subset it fulfills the criteria and gets selected. This is the only selected feature and the final dimension of the feature subset using the RFSA results to seven.

The feature subsets selected through the MFSFSA, FFSA and RFSA are fed as input to the FFSA. The subsets selected through FFSA and RFSA have higher dimension than
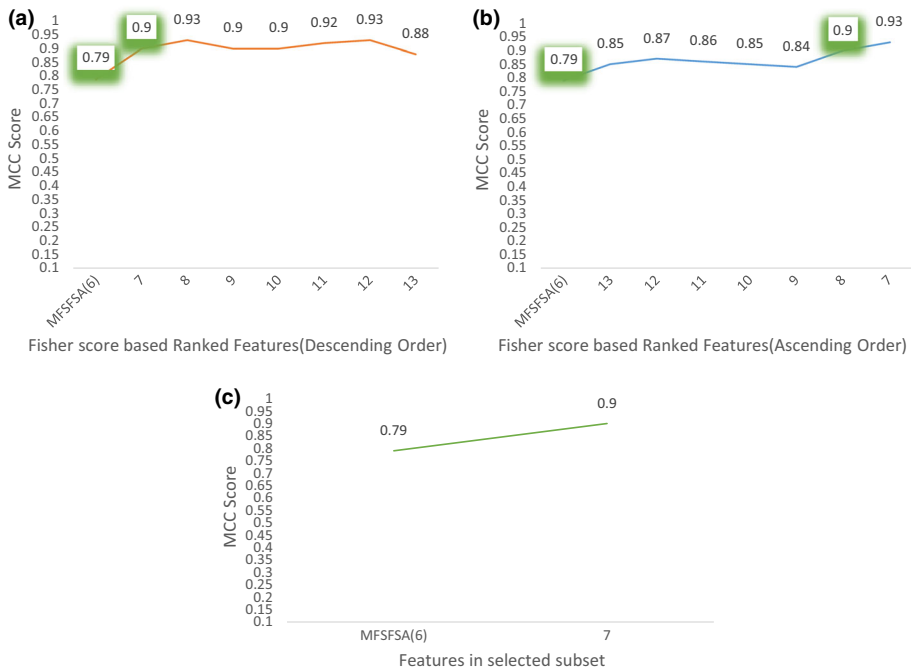
Fig. 3 Feature subset selection using FSSA: Cleveland dataset

of MFSFSA, but higher MCC score as well. The feature subset selected through MFSFSA gets out of the competition. The subsets selected through FFSA and RFSA contain different combinations, but the same MCC score and feature dimension. In this case, as the 7th ranked feature added using the FFSA has a higher Fisher score than the 8th ranked feature added using the RFSA, the former will be used as the final choice. So, the FSSA returns feature subset selected through FFSA and shown in Fig. 3c.

Figure 4 presents a graphical comparative analysis of the proposed technique versus the existing ones when the experiments ate performed over the Cleveland dataset. It may be observed that the proposed technique performs better in terms of accuracy and specificity. In comparison, the highest accuracy was achieved by fuzzy-based CDSS among the existing techniques [15]. In terms of sensitivity, the proposed technique showed moderate results which are close to the highest achieved result, i.e., 76%. In terms of specificity, it was again the proposed technique which performed better than the existing techniques. The closest results in terms of specificity were achieved through the Fuzzy decision tree [22].

5.2 Results for the Hungarian dataset

Figure 5a, b, c shows the results of feature subset selection when the proposed technique is applied to the Hungarian dataset. In all the three graphs, it is obvious that the number of features selected through the MFSFSA is four. These are the highest ranked features as per the Fisher score. The MFS_MCC score achieved through these four selected features is 0.6. Figure 5a represents the MCC score, (i.e, F_MCC), obtained when a feature is added to the feature subset using FFSA. When the 5th ranked feature is added to the feature subset, the F_MCC score for feature subset gets 0.67. This does not fulfill SFSMCC $F\_MCC \geq$ TS+δ_f.
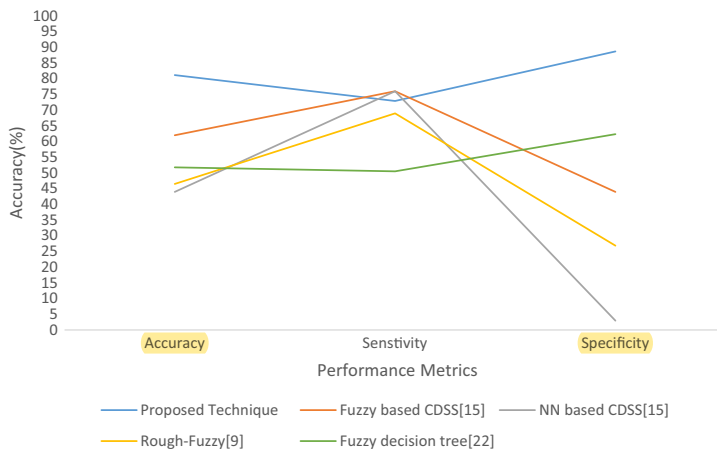
Fig. 4 Comparative analysis: Cleveland dataset

The inclusion of the 5th ranked feature in the selected feature subset boosted the F_MCC score by 0.07, but as per criteria, it should have been at least 0.09. A similar situation arises with the 6th ranked feature and this does not get selected as well. When the 7th ranked feature is added to feature subset, it gets selected as it increases the MCC score by 0.1. Ranked features 8 through 11 do not get selected as they do not boost the MCC score by a minimum score of 0.045. The 12th ranked feature is selected for the feature subset as it increases the MCC score by 0.05. In total, six (6) features are selected with 0.75 as its MCC score.

The same process is performed for the RFSA but with the ranked features inclusion in reverse order. This is shown in Fig. 5b. A subset of 7 features gets selected by attaining a 0.81 MCC score (i.e., R_MCC). Figure 5c shows a graph of the selected features with the R_MCC score achieved by their inclusion. The subsets selected through the MFSFSA, FFSA and RFSA are provided as an input to the FSSA. Through the MFSFSA, the selected subset has four features while subset through RFSA has seven and the FFSA selected a subset with six features. However, dimension of subset selected through RFSA is the highest but achieved same in MCC score as well. So, the subset selected through the RFSA is returned as the final choice.

Figure 6 presents the comparative results of the proposed technique with the existing methodologies in terms of the opted evaluation metrics. In terms of accuracy, it is obvious that the performance of the proposed technique is better than the existing competitors by a considerable margin. Its accuracy is 38% more than the fuzzy-based CDSS technique [15] which had 46% accuracy the highest among its fellows. Likewise, the proposed technique performed better in terms of other two metrics, i.e., sensitivity and specificity, as well.

5.3 Results for the Switzerland dataset

Figure 7 shows the results of the feature subset selection process by applying the proposed technique over Switzerland dataset. In Fig. 7a, a feature subset is selected through the MFS-FSA, FFSA. There are four (04) features which have Fisher scores higher than the mean score and get selected through MFSFSA. The MCC score (MFS_MCC) obtained through the selected features is 0.601. Through the FFSA, when the 5th ranking feature is added to
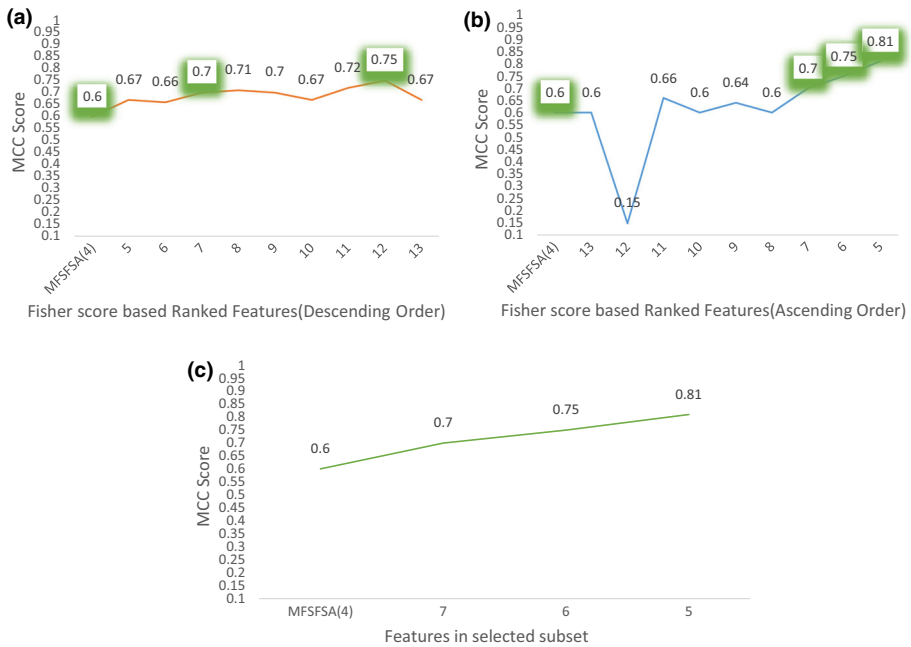
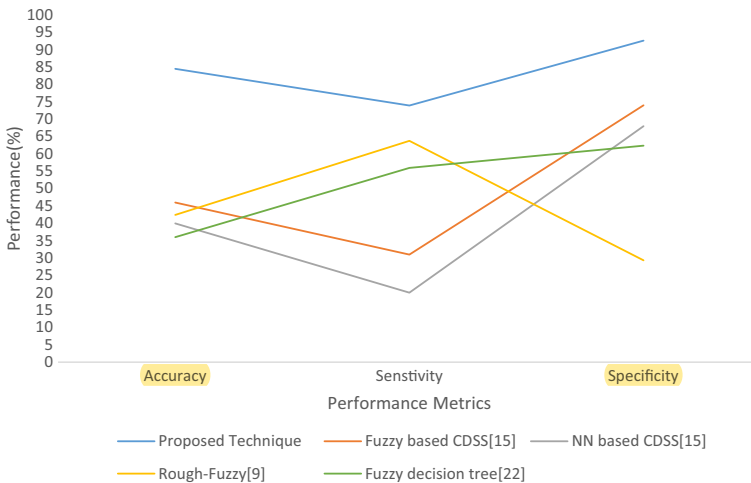**Fig. 5** Feature subset selection using FSSA: Hungarian dataset

Fig. 6 Comparative analysis: Hungarian dataset

the feature subset it attains a 0.70 MCC score. It is selected as it boosted the MCC score (i.e., F_MCC) by 0.1, higher than the minimum required score increase of 0.09. In the next iteration, 6th ranked feature is included in the feature subset and a F_MCC score of 0.759 is achieved. It boosts the F_MCC score by 0.059 score which is more than the required increase of 0.045. The next ranked feature does not affect the F_MCC score, so it is excluded. When the 8th ranked feature is included in the feature subset, the subset achieves 0.92 MCC score
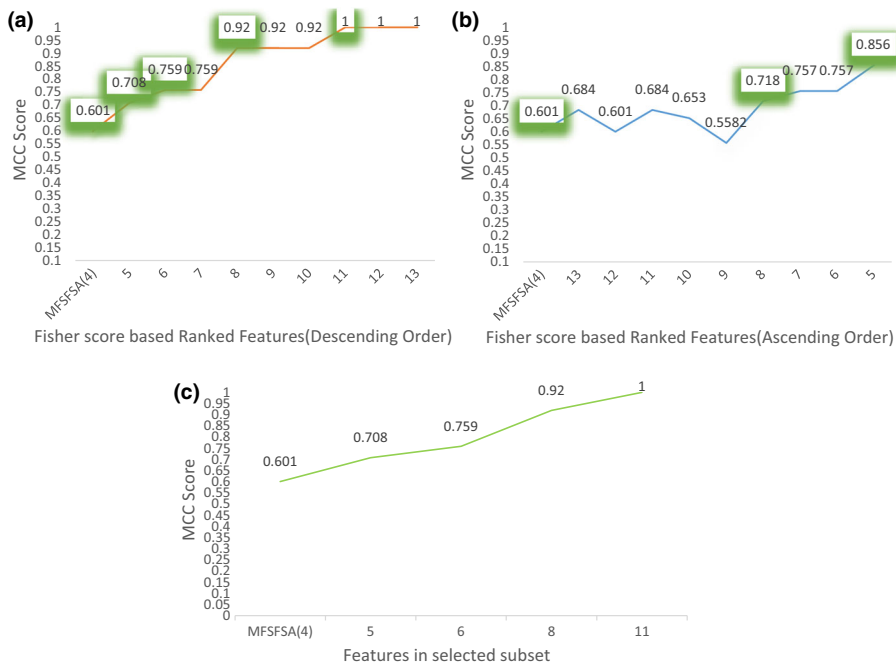
Fig. 7 Feature subset selection using FSSA: Switzerland dataset

and score is boosted by 0.161. This is higher than the required threshold of 0.023, thus allowing it to become part of feature subset. The 9th and 10th ranked features also made no impact over the MCC scores, so they are not selected. The 11th ranked feature, when added to feature subset, the subset achieved an F_MCC score of 1, the maximum possible score. The MCC boost attained is 0.08 greater than the required 0.011. Other features did not make an impact as well. As such, the dimensions of the selected feature subset through the FFSA are eight and F_MCC score of 1 is achieved. Figure 7b represents simulation results of the feature subset selection using the RFSA. As previously mentioned, the Swiss dataset has four features included in the subset through MFSFSA. As features are added in reverse order (i.e., lowest to highest rank), it was determined that the inclusion of features ranked 13th through 9th did not make any considerable impact over the MCC score (R_MCC). When the 8th ranked feature is included, however, it raises the R_MCC score of the feature subset from 0.601 to 0.718. An increase of 0.117, higher than the required raise of 0.09, makes the said feature selected. Features ranked 7th and 6th are also not good choices for the feature subset as their inclusion did not meet the required minimum increase in the R_MCC score. The 5th ranked feature is selected as it boosts the R_MCC score to 0.856, an increase of 0.138 that is higher than the threshold of 0.045. In total, there are six (6) features in the subset whose R_MCC score is 0.856 after the RFSA is done. Finally, FSSA selects feature subset obtained through the FFSA which attained a 1 MCC score. Figure 7c shows MCC scores of the features subset when each selected feature is included.

Figure 8 shows that the comparative performance of the proposed technique in terms of accuracy is better than its fellows. Similarly, the sensitivity score of the proposed technique is quite promising and better than all the comparable techniques while the specificity score was reasonable.
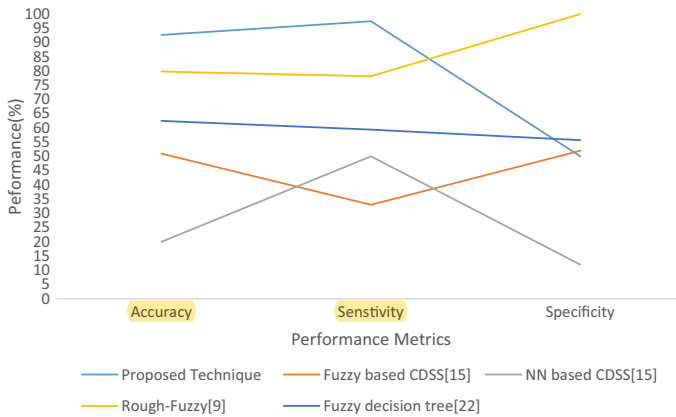
Fig. 8 Comparative analysis: Switzerland dataset

5.4 Results for the SPECTF dataset

The SPECTF dataset is quite different from the other UCI datasets as it has 44-dimensional features. When the proposed technique is applied over SPECTF, 19 are directly selected through the *MFSFSA-NFGMFS* criteria as they have Fisher scores higher than the mean. The MCC score of the selected feature subset already results in 1, the maximum score, so cardiac system algorithm (CSA) decides that there is no need to perform the FFSA and RFSA, thus allowing it to remain as the final selected feature subset. Figure 9 graphically represents the performance based on comparative results of the proposed and eight (8) other existing techniques [5]. A different set of existing techniques were used on the SPECTF dataset due to the number and type of attribute differences that it has over the other datasets (as evident in Tables 1, 2) It can be seen that our proposed technique performed better than the existing ones in terms of its accuracy (82.7%) accurate, except for one (i.e., FireFly+CFARS-AR). In terms of sensitivity and specificity, however, the proposed technique achieved average performance as compared to the others.

5.5 K-fold validations

This section presents the K-fold validations for the 3 UCI datasets and provides a comparison with K-fold validations for the NN-based CDSS [15], Markos [22] and fuzzy-based CDSS [15]. For the Swiss dataset, the eightfold analysis was used, while the tenfold analysis was used for the other two. Each fold contained 30, 28 and 15 records for the Cleveland, Hungarian and the Swiss datasets, respectively. K-fold analysis over the SPECTF is not performed due to the unavailability of the comparative results. The K-fold validations confirm the results achieved by the proposed technique is the most accurate among its fellow techniques. Figures 10, 11, 12 show that the proposed technique has provided better accuracy for almost every single fold. It is important to state that the configurations of the folds used are identical as they were used in all the existing research techniques.
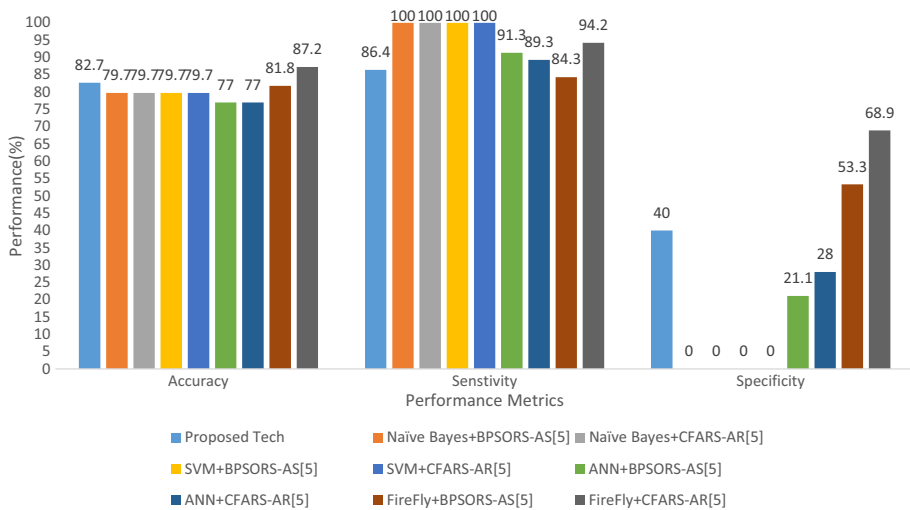
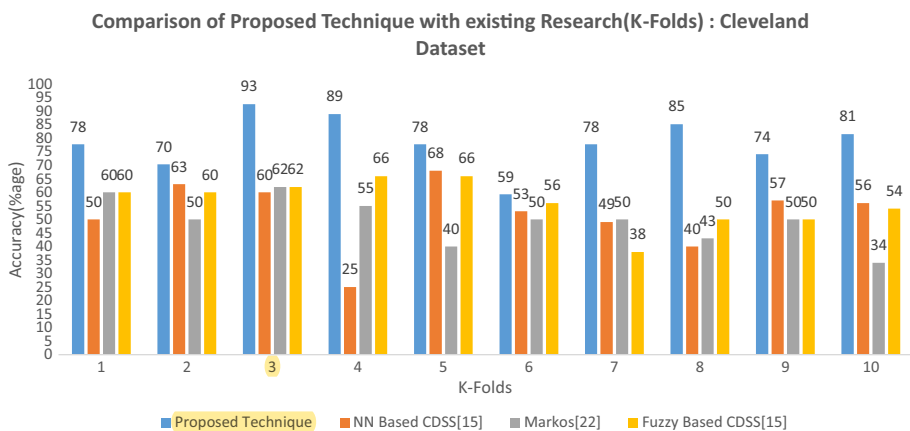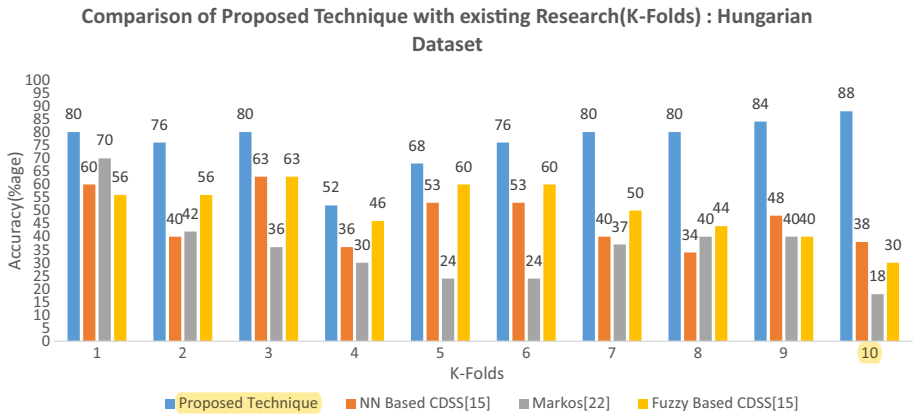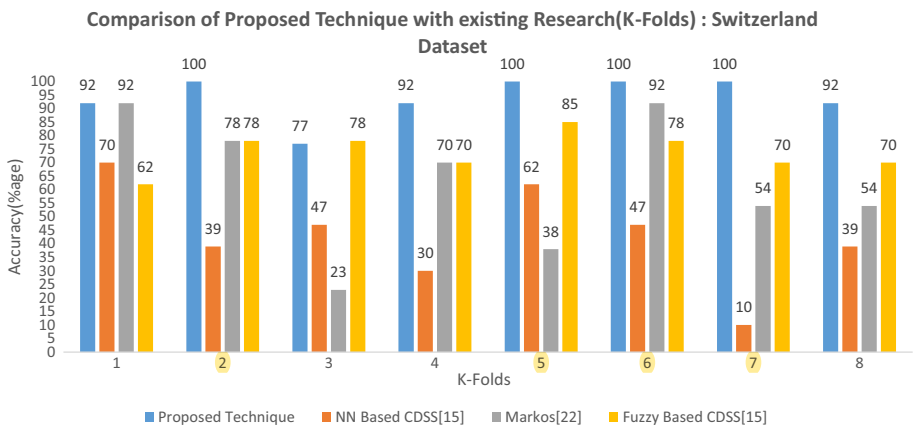**Fig. 9** Comparative analysis: SPECTF dataset

Fig. 10 K-fold comparative analysis: Cleveland

5.6 Computational cost analysis of the proposed technique

The computational complexity of the RBF-SVM is calculated as $O(max(M, d)min(M, d)^2)$, where M represents the number of samples, while d is the feature dimension [31]. As the number of samples are greater than the feature dimension, i.e., $M > d$, the complexity becomes $O(Md^2)$. Using RBF-SVM complexity, the computational cost of the MFSFSA becomes:

$$4 + O(M - T)Q^2) \Rightarrow O((M - T)Q^2) \tag{10}$$

where Q is the number of features selected through MFSFSA. The computational complexity of the FFSA is calculated as:

Fig. 11 K-fold comparative analysis: Hungarian

Fig. 12 K-fold comparative analysis: Switzerland

$$2 + (N - Q)O\left(1 + \left((M - T)\frac{N(N + Q + 1)}{2}\right)^2\right)$$

$$\Rightarrow (N - Q)O\left((M - T)\left(N\left(\frac{N + Q + 1}{2}\right)\right)^2\right) \quad (11)$$

N and T represent original feature dimension and number of instances in test data, respectively. As only training and validations are used in FFSA and RFSA, the test instances are excluded. The RFSA work the same as the FFSA but only chooses the features with Fisher score in reverse order, so its complexity is the same as the FFSA, i.e., $((N - Q)O((M - T)(N(\frac{N+Q+1}{2}))^2))$. The maximum number of statements (including the conditional statements) that come into execution are seven. Once the feature subset of reduced dimension (P) gets selected, the computational cost for the training and testing over the reduced dimensional feature subset is calculated as:

$$O(MP^2) \quad (12)$$

By summing individual complexities of all the algorithms, the overall computational cost of the proposed technique becomes:

$$O((M - T)Q^2) + (N - Q)O\left((M - T)\left(N\frac{(N + Q + 1)}{2}\right)^2\right)$$

$$+ (N - Q)O\left((M - T)\left(N\frac{(N + Q + 1)}{2}\right)^2\right) + 7 + O(MP^2) \tag{13}$$

$$\Rightarrow O((M - T)Q^2) + 2(N - Q)O\left((M - T)\left(N\frac{(N + Q + 1)}{2}\right)^2\right)$$

$$+ (N - Q)O((M - T) + O(MP^2)$$

For the best-case scenario, the complexity is:

$$O((M - T)Q^2) + O(MP^2) \tag{14}$$

This is the case when the features selected through the MFSFSA return maximum MCC, i.e., one. It means not to select the features using FFSA and RFSA as they would not be able to increase the validation MCC score. As in this case P and Q are same, we may rewrite it as:

$$O((M - T)Q^2) + O(MQ^2) \tag{15}$$

The best computational complexity is achieved for the experiments performed over the SPECTF dataset. Out of total 44 features, 19 were selected through MFSFSA and they produced the MCC score of 1. This is the case where selection process through FFSA and RFSA is not performed, resulting the use of the same 19 features for training and testing purpose. For the average cases, the computational complexity remains $O((M - T)Q^2) + 2(N - Q)O((M - T)(N\frac{(N+Q+1)}{2})^2)) + (N - Q)O((M - T) + O(MQ^2)$. For all the other three experiments (Cleveland, Hungarian and Switzerland), the computational complexity remained as average case. These are the cases when all the algorithms went through and $Q > 1$ and $P < N$.

The highest computational complexity occurs for the three cases, i.e., (i) The case when Q is nearly equal to 1. In this scenario, FFSA and RFSA have to go through nearly all the features although P is less than N by a margin. (ii) It is the case when P is greater than 1 by a margin but higher MCC is not obtained, FFSA and RFSA have to go through as average case but P nearly equals N. (iii) The third case is worst of all when Q nearly equals 1 and P nearly equals N. In this case, FFSA and RFSA need to go through all the features and training/testing is performed over the original (approximately) feature dimension.

6 Conclusion

A cardiac disease diagnosis system is presented by proposing the feature selection algorithms, i.e., MFSFSA, FFSA, RFSA and FSSA. The proposed algorithms use individual Fisher scores of the features, along with MCC scores of the subsets, for their selection in the feature subsets. The proposed feature selection algorithms select one feature subset each and among all of them the feature subset with the higher MCC score and lower dimension is a good choice of selection. A binary class RBF kernel-based SVM is used for predicting the heart disease. Four UCI datasets are used for the verification of the proposed technique. The proposed technique is analyzed over three performance metrics, i.e., accuracy, sensitivity

and specificity. The proposed technique achieved considerably better results than the other comparative techniques. A detail of computational analysis for the proposed technique is presented as well.

Compliance with ethical standards

Conflicts of interest There have been no involvements that might raise the question of bias in the work reported or in the conclusions, implications or opinions stated. The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. WHO (2016) Who—cardiovascular diseases (cvds). http://www.who.int/mediacentre/factsheets/fs317/en/. Accessed 10 Jan 2016
2. Naicker S, Plange-Rhule J, Tutt RC, Eastwood JB (2009) Shortage of healthcare workers in developing countries-Africa. Ethn Dis 19(1):60
3. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, Morton SC, Shekelle PG (2006) Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 144(10):742–752
4. Rajkumar A, Reena GS (2010) Diagnosis of heart disease using datamining algorithm. Glob J Comput Sci Technol 10(10):38–43
5. Long CN, Meesad P, Unger H (2015) A highly accurate firefly based algorithm for heart disease prediction. Expert Syst Appl 42(21):8221–8231
6. Ismaeel S, Miri A, Chourishi D (2015) Using the extreme learning machine (elm) technique for heart disease diagnosis. In: 2015 IEEE Canada international on humanitarian technology conference (IHTC2015), pp 1–3
7. Krishnaiah V, Narsimha G, Chandra N (2015) Heart disease prediction system using data mining technique by fuzzy k-nn approach. In: Satapathy SC, Govardhan A, Raju KS, Mandal JK (eds) Emerging ICT for bridging the future—proceedings of the 49th annual convention of the computer society of India (CSI) volume 1, ser. Advances in intelligent systems and computing, vol 337. Springer International Publishing, pp 371–384
8. Chitra R, Seenivasagam V (2015) Heart disease prediction system using intelligent network. In: Power electronics and renewable energy systems. Springer, New York, pp 1377–1384
9. Srinivas K, Rao G, Govardhan A (2014) Rough-fuzzy classifier: a system to predict the heart disease by blending two different set theories. Arab J Sci Eng 39(4):2857–2868
10. Yang GJ, Kim J-K, Kang U-G, Lee Y-H (2014) Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (anfislda). Pers Ubiquitous Comput 18(6):1351–1362
11. Yuehjen ES, Hou C-D, Chiu C-C (2014) Hybrid intelligent modeling schemes for heart disease classification. Appl Soft Comput 14(Part A):47–52, special issue on hybrid intelligent methods for health technologies
12. Muthukaruppan S, Er M (2012) A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. Expert Syst Appl 39(14):11 657–11 665
13. Anooj PK (2012) Clinical decision support system: risk level prediction of heart disease using decision tree fuzzy rules. Int J Res Rev Comput Sci 3(3):1659–1667
14. Anooj PK (2012) Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J King Saud Univ Comput Inf Sci 24(1):27–40
15. Shilaskar S, Ghatol A (2013) Feature selection for medical diagnosis: evaluation for cardiovascular diseases. Expert Syst Appl 40(10):4146–4153
16. Nahar J, Imam T, Tickle KS, Chen Y-PP (2013) Computational intelligence for heart disease diagnosis: a medical knowledge driven approach. Expert Syst Appl 40(1):96–104

17. Giri D, Acharya UR, Martis RJ, Sree SV, Lim T-C, Ahamed T, Suri JS (2013) Automated diagnosis of coronary artery disease affected patients using lda, pca, ica and discrete wavelet transform. Knowl-Based Syst 37:274–282
18. Zhao Z, Wang L, Liu H, Ye J (2013) On similarity preserving feature selection. IEEE Trans Knowl Data Eng 25(3):619–632
19. Hancer E, Xue B, Karaboga D, Zhang M (2015) A binary abc algorithm based on advanced similarity scheme for feature selection. Appl Soft Comput 36:334–348
20. Zhang J, Yu J, Wan J, Zeng Z (2015) l2, 1 norm regularized fisher criterion for optimal feature selection. Neurocomputing 166:455–463
21. Li C, Shi C, Zhang H, Hui C, Lam K-M, Zhang S (2014) Cost-sensitive feature selection in medical data analysis with trace ratio criterion. In: 2014 12th international conference on signal processing (ICSP). IEEE, pp 1077–1082
22. Markos GT, Exarchos TP, Fotiadis D, Kotsia AP, Vakalis KV, Naka KK, Michalis LK et al (2008) Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling. IEEE Trans Inf Technol Biomed 12(4):447–458
23. Gu Q, Li Z, Han J (2012) Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725
24. Lee C, Lee GG (2006) Information gain and divergence-based feature selection for machine learning-based text categorization. Inf Process Manag 42(1):155–165
25. Shah SMS, Batool S, Khan I, Ashraf MU, Abbas SH, Hussain SA (2017) Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis. Phys A Stat Mech Appl 482:796–807
26. Akay MF (2009) Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 36(2):3240–3247
27. Çomak E, Arslan A, Türkoğlu İ (2007) A decision support system based on support vector machines for diagnosis of the heart valve diseases. Comput Biol Med 37(1):21–27
28. Yahiaoui A, Er O, Yumusak N (2017) A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines. Biomed Res 28(9):4208–4212
29. Huang M-W, Chen C-W, Lin W-C, Ke S-W, Tsai C-F (2017) SVM and SVM ensembles in breast cancer prediction. PloS one 12(1):e0161501
30. Asuncion A, Newman D (2007) UCI machine learning repository. http://www.ics.uci.edu/~mlearn/MLRepository.html. Accessed 15 June 2015
31. Chapelle O (2007) Training a support vector machine in the primal. Neural comput 19(5):1155–1178

Syed Muhammad Saqlain is working as Assistant Professor at the Department of Computer Science & Software Engineering, International Islamic University Islamabad, Pakistan. He has more than 15 years of experience in research, development and teaching. His research interests include image processing, computer vision, activity analysis, sign language recognition, pattern classification and data mining.

Muhammad Sher is currently working as Dean of Faculty of Basic & Applied Sciences, International Islamic University Islamabad. He has been teaching at the Department of Computer Science & Software Engineering since 1996 and is now a Professor. He also has served as the Chairman of Computer Science & Software Engineering from 2008 to 2014. He is also the Director IT, International Islamic University, Islamabad. He has been actively involved in research and supervised many successful MS and PhD graduates. He received his Ph.D. in computer science from TU Berlin, Germany, and M.Sc. from Quid-e-Azam University, Islamabad. He authored more than 100 publications in different conferences and journals of international repute. His research interests include next-generation networks, IP multimedia sub-systems and network security, cryptography and wireless communication.

Faiz Ali Shah is a PhD scholar at Institute of Computer Science, University of Tartu, Estonia. He completed his MS computer Science in 2011. He has an experience of more than 12 years in software industry and academia. His areas of interest are pattern data mining, image processing and software engineering.

Imran Khan received his MCS, MS and Ph.D Computer Science Degree from International Islamic University Islamabad, Pakistan, in 2002, 2005 and 2017, respectively. Currently he is an Assistant at the Department of Computer Science & Software Engineering. His research interests include information security and privacy, exchange and medical data management, IP multimedia sub-system, next-generation networks and cloud computing.

Muhammad Usman Ashraf is a Ph.D. scholar at the department of Computer Science & Software Engineering, International Islamic University Islamabad, Pakistan. He is also serving as a Lecturer at the IBMS, University of Agriculture Faisalabad, Pakistan. His multidisciplinary interests include HRI, distributed healthcare systems and network security.

Muhammad Awais completed his PhD from University of Bayreuth, Germany in 2013. Currently he is serving as an Assistant Professor at Government University College Faisalabad, Pakistan. His area of interest are human computer interaction/ collaboration and data mining.

Anwar Ghani received his Ph.D. Computer Science degree in 2016 and MS Computer degree in 2011 from International Islamic University, Islamabad. During Ph.D. and MS, he received exchange scholarship from European Commission to study at VU University Amsterdam and Barno University Czech Republic. He received his Bachelor degree in Computer Science from the University of Malakand K.P.K, Pakistan in 2007. Currently, he is working as Lecturer at the Department of Computer Science & Software Engineering, International Islamic University, Islamabad. He is also served as a researcher in ICT R&D project at International Islamic University. His broad research interests include next-generation networks, information security, wireless sensor networks, energy-efficient collaborative communication.