IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 May 2022, accepted 14 June 2022, date of publication 21 June 2022, date of current version 29 June 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3185129

Improving the Heart Disease Detection and Patients' Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest

ABDALLAH ABDELLATIF[1], (Member, IEEE), HAMDAN ABDELLATEF[2], (Member, IEEE), JEEVAN KANESAN[1], CHEE-ONN CHOW[1], (Senior Member, IEEE), JOON HUANG CHUAH[1], (Senior Member, IEEE), AND HASSAN MUWAFAQ GHENI[3]

[1]Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[2]Electrical and Computer Engineering Department, School of Engineering, Lebanese American University, Byblos, Lebanon
[3]Computer Techniques Engineering Department, Al-Mustaqbal University College, Hillah 51001, Iraq

Corresponding author: Jeevan Kanesan (jievan@um.edu.my)

This work was supported by the Ministry of Higher Education through the Fundamental Research Grant Scheme under Grant FRGS/1/2020/ICT02/UM/02/2.

ABSTRACT Heart disease is the leading cause of death worldwide. A Machine Learning (ML) system can detect heart disease in the early stages to mitigate mortality rates based on clinical data. However, the class imbalance and high dimensionality issues have been a persistent challenge in ML, preventing accurate predictive data analysis in many real-world applications, including heart disease detection. In this regard, this work proposes a new method to address these issues and improve the predict the presence of heart disease and patients' survival, including supervised infinite feature selection (Inf-FS_S) to find the most significant features and Improved Weighted Random Forest (IWRF) to predict heart disease, and Bayesian optimization to tune the new hyperparameters for IWRF. Two public datasets, including Statlog and heart disease clinical records, were used to develop and validate the proposed model. The proposed model is compared with other hybrid models to show its superiority using performance metrics like accuracy and f-measure to evaluate the models' performance. The results have shown that the proposed Inf-FS_S-IWRF achieved better results than other models in attaining higher accuracy and F-measure on both datasets. Additionally, a comparative study has been performed to compare with previous studies, where the proposed model outperformed the others by an accuracy improvement of 2.4% and 4.6% on both datasets, respectively.

INDEX TERMS CVD detection, heart disease classification, feature selection, random forest, imbalance, Bayesian optimization.

I. INTRODUCTION

Cardiovascular disease (CVD), often known as heart disease, is the leading cause of mortality worldwide. According to recent research conducted by the World Heart Federation, one in every three deaths is caused by cardiovascular disease [1]. By 2030, the World Health Organization (WHO) estimates that over 23.6 million people will die from CVD, primarily heart failure and strokes [2]. Preventing CVD at an early stage is the only approach to halt this kind of mortality and reduce the overall death count. It is relatively difficult to diagnose CVD due to many contributing variables such as high cholesterol, high blood pressure, smoking, diabetes, overweight and obesity, and many other factors. Researchers have been testing various strategies to detect CVD. However, disease prediction at an early stage is difficult due to multiple constraints, including but not limited to method complexity, feature selection, and execution time [3]. As a result, developing effective detection and prediction methods may save countless lives.

Clinical decision-making based on hybrid machine learning (ML) models is used in the medical field to achieve good results. However, the clinical datasets have some challenges, mainly due to the high dimensionality and the class

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

imbalance. As a result, applying ML without addressing these issues affects the approaches' accuracy [4]. Therefore, various ML-based predicting methods for CVD detection and survival have been proposed in the literature. Earlier studies used different ML algorithms to detect CVD, focusing on feature selection (FS). Including rough sets (RS) to select the most significant features and feed them to the chaos firefly algorithm [5] and backpropagation neural network (BPNN) [6] to predict heart disease (HD). In addition, Amin *et al.* and Chicco detect CVD by employing vote with Naive Bayes (NB) and logistic regression (LR) on selected features to predict HD presence and patients' survival [7], [8]. The authors in [9] conducted a comparative analysis study. The authors employed different classifiers on various datasets. They concluded that the conditional inference tree forest (cforest) surpassed the other classifiers.

Haq *et al.* conducted a comparative analysis on a hybrid model constructed using various feature selection strategies and machine learning models. Their research established that reducing features affected the models' performance. According to the study, a combination of Relief-LR delivers maximum accuracy [10]. Gupta *et al.* created a framework for machine intelligence that includes factor analysis of mixed data (FAMD) and random forest (RF). The FAMD was used to identify significant characteristics, and the RF was used to predict CVD [11]. Khan and Algarni [12] developed an Internet of Medical Things (IoMT) to predict HD. The developed model used Modified Slap Swarm Optimization (MSSO) to optimize the adaptive neuro-fuzzy inference system (ANFIS) parameters.

Also, Ali *et al.* proposed two stacked SVMs predicting CVD presence. The first SVM removed the non-significant attributes, while the second SVM was utilized to indicate CVD presence and absence of the model tuned using a hybrid grid search algorithm [13]. Tama *et al.* [14] developed a two-tier ensemble model to detect CVD. The model stacked architecture is intended to combine the CVD forecast of the selected ensemble learners XGBoost, RF, and gradient boosting machine (GBM). In addition, the authors employed particle swarm optimization to choose the most significant features. The authors in [15], [16] developed an IoT framework to evaluate the CVD status. The developed model employed a modified deep convolution neural network (MDCNN) to predict the patient's status based on data received from the sensor. However, since the clinical datasets are imbalanced, the above studies have some limitations in detecting CVD through the mentioned methods.

Some studies have developed reliable CVD detection and patient survival models to address this issue. For example, Ishaq *et al.* applied the synthetic minority over-sampling technique (SMOTE) to balance data distribution and extremely randomized trees (ET) on selected attributes using random forest importance ranking to predict the patients' survival [17]. In addition, Fitriyani *et al.* developed a hybrid method to detect HD consisting of density-based spatial clustering applications with noise (DBSCAN) to detect and remove outliers instances applied to the features selected from information gain. Then, hybrid SMOTE-ENN was employed to balance the dataset and extreme gradient boosting (XGBoost) for CVD detection [18]. Recently, Waqar *et al.* proposed SMOTE-based deep learning to predict heart disease. The authors applied SMOTE technique to balance the dataset without the need for feature selection [19]. However, the balancing method SMOTE has limitations, including blindness of neighbour selection, instance overlapping, small disjuncts, and noise interference [20]–[22]. The related work is summarized in Table 1 regarding the method utilized, feature selection, data balancing, validation method and the datasets used.

Based on prior research, there is still a lack of a model to address the imbalance issue on an algorithm level instead of a data level to improve the accuracy of CVD detection and survival. Therefore, we proposed an Improved Weighted Random Forest (IWRF) to address the imbalanced dataset classification based on cost-sensitive learning to cope with those limitations. Moreover, we integrate the proposed IWRF with supervised infinite feature selection (Inf-FS$_S$) for feature ranking and selection and Bayesian Optimization (BO) to optimize the IWRF weighting coefficient. Prior studies have reported that the model prediction performance significantly improved by integrating Inf-FS [23]–[25] and optimizing the ML model using BO [26], [27]. However, to the best of our knowledge, no studies integrated Inf-FS$_S$ and BO with IWRF to predict the presence and survival of CVD.

Therefore, we propose an effective method to predict CVD and patients' survival: Inf-FS$_S$ to rank the features by importance and select the best features, IWRF to predict CVD, and BO to find the best weighting coefficient. Two public datasets were chosen to develop the model and test the model, the Statlog dataset [28] to detect the absence and presence of CVD and the heart failure clinical record dataset [29] to predict the patients' survival. So, we set out to develop ML algorithms to diagnose CVD and patient survival to assist healthcare professionals. As a result, early treatment might be implemented to avoid the deaths caused by late CVD detection. The main contributions of this study can be summarized as follows:

- An Improved Weighted Random Forest (IWRF) is developed to deal with class imbalance.
- Decision support (Inf-FS$_S$-BO-IWRF) is proposed to predict the presence of CVD and patients' survival.
- The evaluation of the proposed IWRF model in comparison to other ML models such as SVM, kNN, XGBoost, and SMOTE-RF highlights the superiority of the proposed model
- Identify the most important attributes in the dataset that impact the machine learning system performance.
- The effectiveness of the proposed Inf-FS$_S$-BO-IWRF is evaluated on two binary public datasets.

The rest of the paper is structured as follows: Section 2 presents the related studies. Section 3 is the proposed methodology. Section 4 describes the performance evaluation metrics, results and discussion for the conducted experiments, and the state-of-art comparison. Finally, the conclusion and future work are presented in section 5.

II. RELATED STUDIES

Weighted ensembles have been the subject of a wide range of research investigations. Pham *et al.* developed a weighted approach for generalizing bootstrap-aggregated ensemble learning to a weighted vote by evaluating various averaging methods [30]. Later, Pham *et al.* proposed a Cesaro average-based to enhance the RF for the binary classification issue. This strategy is driven by the inherent instability of tree-based prediction averaging [31]. Next, Chen *et al.* introduced the weighted random forest (WRF), incorporating cost-sensitive learning. It weights both the majority and minority instances in a training set, with a higher weight for minority instances. Also, Chen pioneered the balanced random forest (BRF) approach, which involves sampling. BRF was developed to account for the likelihood that some bootstrap samples generated will include fewer or no minority cases. The core concept of BRF is to generate bootstrap samples through systematic under-sampling of the majority class [32].

To detect credit card fraud, Xuan *et al.* developed Refined Weighted RF (RWRF). The enhancement is in two areas. They utilized all training data (both Out-of-Bag (OOB) and In-Bag (INB) data) because they believed that evaluating the performance of various base classifiers should be done using the same dataset. Additionally, they utilized the gap between the chance of correctly predicting true and false class labels to determine how the predicted number of votes for the correct label surpasses the anticipated number of votes for the incorrect label [33]. Kulkarni *et al.* discussed efforts to increase the accuracy and time required for training the RF classifier. They are based on disjoint partitioning of training datasets, the usage of split measures or multiple feature evaluation to generate RF base decision trees, the use of weighted voting rather than majority voting, the usage of diversity in bootstrap datasets to create the most diverse classifiers, and the usage of dynamic programming method to discover the best subset of RF [34].

A probabilistic approach for combining classifiers was presented by Kuncheva *et al.* The four combination approaches, including recall combiner, majority vote, I Bayes combiner, and weighted majority vote. It provides strict optimality requirements (lowest classification error) for each. Both the class-conditional independence of classifier outcomes and the presumption of certain accuracy form the basis of the framework [35]. Gajowniczek *et al.* proposed a new weighting approach with tunable parameters that apply to each RF tree [36]. The classification strategy of hybrid NB and sample weighted RF (SWRF) used by Babu *et al.* for sub-acute ischemic stroke lesion segmentation was a successful meta-heuristic feature selection method. NB is taught and used to estimate training sample weights in this example. To train SWRF, a set of training samples with predicted weights is used [37].

TABLE 1. A chronological overview of existing systems for heart disease diagnosis and patient survival.

Study	Method	FS	Data Balance	Validation	Dataset
[5]	CFARS-AR	Yes	No	No	Statlog, SPECTF
[6]	RS-BPNN	Yes	No	10 CV	Statlog
[39]	LR	No	No	10 CV	Statlog
[10]	Relief-LR	Yes	No	10 CV	Cleveland
[7]	Vote with NB & LR	Yes	No	10 CV	Cleveland, Statlog
[13]	Stacked SVM	Yes	No	Holdout	Cleveland
[11]	FAMD-RF	Yes	No	Holdout	Cleveland
[15]	MDCNN	Yes	No	10 CV	Cleveland, Framingham , Public Health
[14]	Two-tier ensemble PSO based feature selection	Yes	No	10 CV	Statlog, Z-Alizadeh Sani, Cleveland, Hungarian
[8]	LR	Yes	No	Holdout	HD clinical record
[9]	CF	No	No	10 CV	Statlog
[12]	MSSO-ANFIS	Yes	No	Not mentioned	Cleveland
[18]	DBSCAN + SMOTE-ENN + XGBoost	Yes	Yes	10 CV	Statlog, Cleveland
[17]	SMOTE-RF-ET	Yes	Yes	10 CV	HD clinical record
[19]	SOMTE-ANN	No	Yes	Not mentioned	Cleveland
Current Study	Inf-FS$_s$-BO-IWRF	Yes	Yes	10 CV	Statlog, HD clinical record

Recently, Utkin *et al.* presented a weighted Random Survival Forest (RSF) as a way to improve the performance of the RF. The suggested model's core idea is to replace the traditional averaging approach used to estimate the RSF hazard function with weighted averaging. Each tree is given a set of weights, which can be considered training parameters. They are calculated by solving a conventional quadratic optimization problem to maximize Harrell's C-index [38]. Bader-El-Den *et al.* introduced Biased Random Forest (BRF). Rather than boosting minority occurrences in data sets, BRF tries to oversample the classification ensemble by expanding the number of classifiers that represent the minority class in the ensemble. The BRF technique uses the kNN algorithm to determine the crucial regions within a data set. The conventional RF is supplemented with additional random trees based on the key locations [21].

III. PROPOSED METHODOLOGY

The proposed method is developed to obtain a high-performance heart disease prediction of the presence and patients' survival of CVD. Figure 1 presents the flowcharts of the proposed method.

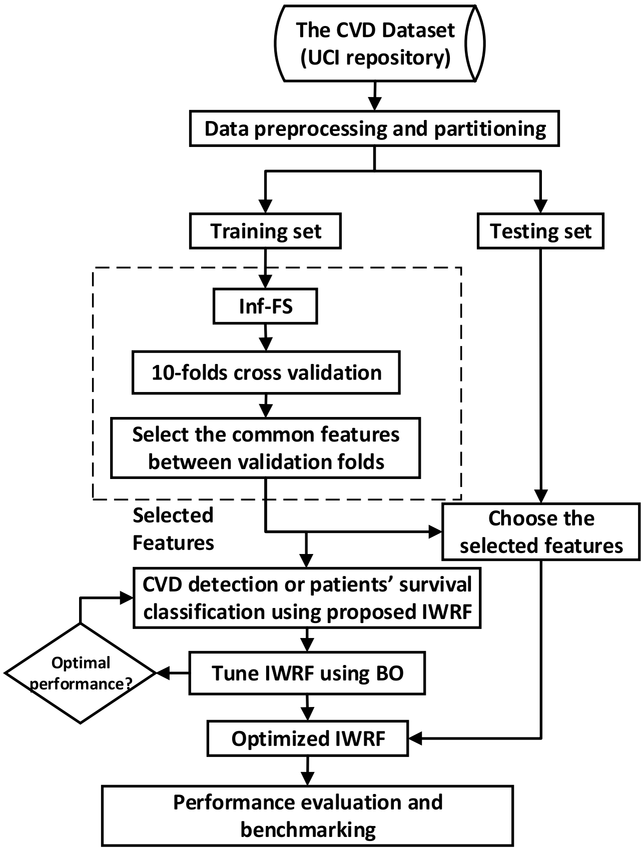

The CVD Dataset (UCI repository)
Data preprocessing and partitioning
Training set
Testing set
Inf-FS
10-folds cross validation
Select the common features between validation folds
Selected Features
Choose the selected features
CVD detection or patients' survival classification using proposed IWRF
Optimal performance?
Tune IWRF using BO
Optimized IWRF
Performance evaluation and benchmarking

FIGURE 1. The proposed flowchart for heart disease detection and patients' survival.

A. FEATURE SELECTION USING INFINITE FEATURE SELECTION

Feature selection is critical in ML since the performance of ML techniques is strongly reliant on the features selected. Various features can obscure and entangle the data's different explanatory components [40]. There are multiple approaches for choosing the best attributes in the literature. We have selected a recently developed FS approach called infinite feature selection supervised (Inf-FS$_s$) [41] for this study. This approach is graph-based feature filtering that makes the ranking by considering all the potential subsets of features work in a supervised and unsupervised form. It is constructed upon a fully connected weighted undirected graph G = (V, E). The nodes V denote all features, and the edges E reflect the pairwise relationships between them. Consider G as an adjacency matrix A, where each of its components a_{ij} ($1 \leq I, j \leq n$) represents the degree of confidence that the nodes $\vec{v}_i$ and $\vec{v}_j$ are both potential candidates for selection done with the following weight function:

$$\mathrm{A}\,(I,j) = \varphi\left(\vec{v}_i, \vec{v}_j\right) \tag{1}$$

where weight function φ is real-valued, that specifies the value of each edge. In Inf-FS$_s$, the weight function integrates class labels utilizing Fisher criteria and mutual information. Therefore, the weight function $\varphi(\vec{v}_i, \vec{v}_j)$ is produced by three factors, Fisher criteria (h_i), normalized mutual information (m_i), and normalized standard deviation (σ_i). The following equations calculate the factors:

$$h_i = \frac{|\mu_{i,1} - \mu_{i,2}|^2}{\sigma_{i,1}^2 + \sigma_{i,2}^2} \tag{2}$$

$$m_i = \sum_{y \in Y} \sum_{z \in f_i} p\,(z, y) \log\left(\frac{p(z,y)}{p\,(z)\,p(y)}\right) \tag{3}$$

where $\sigma_{\mathrm{i,g}}$, and $\mu_{\mathrm{i,g}}$ represent the standard deviation and mean for ith attributes considering the instances of gth class. The ith feature is less redundant the closer h_i is to 1 since it doesn't overlap with the other domain. While Y and $p(z, y)$ are the class labels and joint probability distribution, respectively. In practice, m_i is a measure of how much a feature vector's knowledge reduces the level of uncertainty about each class. Also, the normalized standard deviation (σ_i) is normalized to a range of [0, 1] by the maximum std over the set features (F). Finally, the three-element are weighted linearly.

$$s_i = h_i\alpha_1 + m_i\alpha_2 + \sigma_i\alpha_3 \tag{4}$$

The parameters α_{k} is the mixing coefficients where α_{k} belongs to a range of [0,1], $\sum_k \alpha_k = 1$. Their values are set during experiments. The score s_i shows how much a feature is not redundant and relevant to other classes. Finally, the adjacency matrix A's weights are constructed by coupling the correspondent s in the following manner:

$$\mathrm{A}(i,j) = \varphi(\vec{v}_i, \vec{v}_j) = s_i s_j \tag{5}$$

After the adjacency matrix is constructed, ranking is performed while evaluating the redundancy of the features, taking into account all possible pathways among the nodes. The Inf-FS$_s$ algorithm is described in detail [41]. Finally, cross-validation (cv) is used to determine the mixing coefficient α_k for each training split of both datasets.

B. IMPROVED WEIGHTED RANDOM FOREST

Bagged (bootstrap-aggregated) DTs can reduce overfitting effects and improve generalization by merging the outcomes of several DTs. In the bootstrap aggregating learning concept, T base models (decision trees) are trained over subgroups taken with replacement from the dataset. Their results are voted to create a prediction estimate of the model. Voting and bagging are implemented to reduce the model's variance without raising its bias since base models are provided with multiple training sets, creating a varied ensemble [25]. A bootstrap of M' samples is picked randomly from the initial M training samples and replaced for every tree t, where t belongs [1, T]. During the training process, $F' < F$ attributes are randomly chosen from F available features at each tree node, and the optimal split is determined by applying those F' attributes. In the testing process, the unseen instance is run through all the T trees in the forest, resulting in T predictions for the test instance. Finally, these forecasts are pooled via voting to provide the final prediction.

In imbalanced data classification, RF classifiers tend to be biased in the direction of the major class since standard RF treats both classes equally. However, several studies have shown that a weighted RF can deliver better prediction results. For this reason, this study presents an Improved Weighted Random Forest (IWRF), which assigns a weight for each class, a higher weight for the minor class. The class weight in the random forest can be computed using the inversely proportional class frequencies in the training dataset. The class weights are presented as the following:

$$CW_1 = \frac{M}{2M_1} \ \& \ CW_2 = \frac{M}{2M_2} \tag{6}$$

where M presents the total number of samples in the dataset, M_1 and M_2 show the number in major and minor classes. We assign a new coefficient, the weighting factor (α), to compute class weights. Thus, the class weights will be calculated as follows:

$$CW_1 = \alpha_1 \frac{M}{2M_1} \ \& \ CW_2 = \alpha_2 \frac{M}{2M_2} \tag{7}$$

where α_1 and α_2 are the weighting factor for major and minor classes, respectively, α_1 and α_2 vary in a range from [0, 1] with default values M_1/M_2 and one for α_1 and α_2. To ensure that CW_2 is always greater than CW_1 to have a heavier penalty on misclassifying the minor class, the weighting factor is subjected to the constrain as follows:

$$\frac{\alpha_1}{M_1} < \frac{\alpha_2}{M_2} \tag{8}$$

The RF algorithm incorporates class weights in two places. Class weights are used in the tree induction technique to weight the Gini criteria for detecting splits. Class weights are again considered at each tree's terminal nodes. Each terminal node's class prediction is established by a weighted majority vote. Moreover, in imbalance classification, there is a substantial chance that a bootstrap sample has few or no instances of the minority class, leading to a tree with low performance in predicting the minority class. A new coefficient (p) is added to control the number of minor class samples in each bootstrap to overcome this problem. It randomly draws a bootstrap from both classes, containing at least one-third of the minority class out of the total samples in the bootstrap. The value of p can range between ($1/3 \le p < 1/2$) depending on the imbalance ratio between the majority and minority classes.

TABLE 2. **Feature ranking and weight importance of a particular fold determined by Inf-FS$_S$.**

Statlog dataset			Heart disease clinical record		
Attributes	Rank	Weight	Attributes	Rank	Weight
Age	11	6.133	Age	7	7.203
Gender	3	9.928	Anaemia	1	11.191
CP	6	8.898	CPK	11	6.403
Tresthps	12	5.793	Diabetes	2	11.135
Chol	13	5.619	Ejection_fraction	8	6.849
Fbs	10	7.159	High BP	3	10.768
Restecg	4	9.875	Platelets	12	6.382
Thalach	8	7.758	Serum_creatinine	10	6.404
Exang	2	11.092	Serum_sodium	9	6.728
Oldpeak	9	7.722	Gender	4	10.766
Slope	7	7.897	Smoking	5	10.587
Ca	5	9.231	Time	6	8.04
Thal	1	13.165			

TABLE 3. **Selected features of both datasets.**

Dataset	Selected features
Statlog	Thal, Exang, Gender, Restecg, Ca, CP, Slope, Thalach, Oldpeak
HD clinical record	Anemia, Diabetes, High BP, Gender, Smoking, Time, Age, Ejection_fraction, Serum_sodium

C. BAYESIAN OPTIMIZATION

BO [42] is a reiterative algorithm widely used for HPO problems. BO applies two key components to define hyperparameter configuration: an acquisition function and a surrogate model [30]. The surrogate model seeks to fit all the examined observations into the objective function. After finding the probabilistic surrogate model's predictive distribution, the acquisition function determines various points by balancing exploitation and exploration. Exploration tests the samples in the areas that have not been tested. In contrast, exploitation tests in the currently promising regions where the global optimum is most expected to occur, depending on the posterior distribution. Bayesian tuning models balance exploitation and exploration processes to determine the current most expected optimal regions and avoid losing better configurations in the unexplored regions [43]. After each evaluation of the objective function, the surrogate model is updated. BO models are sequential processes that are difficult to parallelize since they are built on previously tested variables. Still, within a few iterations, BO can find nearby optimal hyperparameter coefficients [44].

BO's common surrogate model is the tree-structured Parzen Estimator (TPE) [45]. BO-TPE creates two generative models for all domain variables, g(x) and l(x). [46]. The observations are divided into poor and good results by a specified percentile y*; both sets are modeled by simple Parzen windows [45]:

$$p(x|y, D) = \begin{cases} l(x) & y < y^* \\ g(x), & y > y^* \end{cases} \tag{9}$$

D is the search space of the hyperparameter. After that, the acquisition function's expected improvement is reflected by the ratio between $g(x)$ and $l(x)$, which is applied to establish the latest configurations for evaluation. The PE is created in a tree structure to ensure that the necessary conditional

TABLE 4. Performance evaluation of Statlog dataset.

Model	Without FS						With FS					
	Acc.	Pre.	Recall	F-measure	SPC	MCC	Acc.	Pre.	Recall	F-measure	SPC	MCC
SVC	0.921	0.947	0.847	0.894	0.969	0.836	0.929	0.948	0.866	0.905	0.969	0.852
kNN	0.87	0.886	0.771	0.821	0.933	0.728	0.895	0.914	0.809	0.857	0.951	0.781
G-NB	0.907	0.944	0.81	0.872	0.97	0.806	0.907	0.9	0.857	0.878	0.939	0.804
LR	0.903	0.916	0.828	0.87	0.951	0.797	0.907	0.917	0.838	0.875	0.951	0.804
XGBoost	0.933	0.939	0.885	0.91	0.963	0.859	0.944	0.949	0.904	0.926	0.969	0.882
RF	0.929	0.938	0.876	0.905	0.963	0.851	0.94	0.949	0.895	0.92	0.969	0.875
IWRF	0.955	0.98	0.904	0.94	0.987	0.906	**0.977**	**0.963**	**0.98**	**0.97**	**0.975**	**0.954**

TABLE 5. Performance evaluation of HD clinical records dataset.

Model	Without FS						With FS					
	Acc.	Pre.	Recall	F-measure	SPC	MCC	Acc.	Pre.	Recall	F-measure	SPC	MCC
SVC	0.839	0.762	0.677	0.716	0.909	0.609	0.849	0.771	0.711	0.74	0.909	0.636
kNN	0.786	0.699	0.511	0.576	0.904	0.459	0.809	0.712	0.622	0.663	0.89	0.535
G-NB	0.833	0.75	0.667	0.706	0.905	0.592	0.85	0.765	0.722	0.743	0.9	0.63
LR	0.843	0.759	0.7	0.728	0.905	0.619	0.839	0.756	0.689	0.72	0.9	0.61
XGBoost	0.889	0.804	0.801	0.802	0.926	0.726	0.912	0.827	0.83	0.827	0.942	0.769
RF	0.893	0.813	0.801	0.806	0.93	0.733	0.909	0.817	0.83	0.823	0.937	0.762
IWRF	0.933	0.851	0.871	0.86	0.952	0.814	**0.959**	**0.926**	**0.9**	**0.913**	**0.978**	**0.881**

TABLE 6. Comparison results between IWRF and SMOTE-RF on Statlog dataset.

Model	Without Optimization						With Optimization					
	Acc.	Pre.	Recall	F-measure	SPC	MCC	Acc.	Pre.	Recall	F-measure	SPC	MCC
SMOTE-RF	0.947	0.952	0.912	0.93	0.97	0.891	0.978	0.979	0.965	0.972	0.986	0.955
IWRF	0.977	0.963	0.98	0.97	0.975	0.954	**0.983**	**0.986**	**0.972**	**0.979**	**0.991**	**0.966**

TABLE 7. Comparison results between IWRF and SMOTE-RF on HD clinical record dataset.

Model	Without Optimization						With Optimization					
	Acc.	Pre.	Recall	F-measure	SPC	MCC	Acc.	Pre.	Recall	F-measure	SPC	MCC
SMOTE-RF	0.939	0.866	0.883	0.872	0.956	0.831	0.962	0.922	0.919	0.919	0.975	0.892
IWRF	0.959	0.926	0.9	0.913	0.978	0.881	**0.972**	**0.944**	**0.943**	**0.943**	**0.982**	**0.922**

dependencies are maintained. Thus, TPE adopts specific conditional hyper-parameters naturally [44].

IV. EXPERIMENTAL RESULTS

This section presents the feature selection results first, followed by HD presence and survival classification performed for both datasets. The developed model was built and tested for HD on Statlog and heart failure clinical record datasets. The Statlog dataset consists of 14 attributes with the status label, 270 cases, 150 for HD absence and 120 for HD presence. The heart failure clinical record dataset consists of 13 attributes with the survival label, 299 total cases, 202 patients survive, and 97 patients deceased. We used a 10-fold cv procedure in our experiment to avoid overfitting [47]. We evaluated the proposed model using six performance metrics. The confusion matrix is applied to measure the model's output, including True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP). The six-performance metrics are calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{10}$$

TABLE 8. Performance evaluation of proposed method compared with previous studies for HD detection using Statlog dataset.

Author	Method	Performance Evaluation					
		Accuracy	Precision	Recall	Specificity	F_1	MCC
Long et al. [5]	CFARS-AR	0.883		0.849			
Nahato et al.[6]	RS-BPNN	0.904		0.946			
Dwivedi [39]	LR	0.85	0.85	0.89		0.87	
Amin et al. [7]	Vote with NB + LR	0.874					0.851
Tama et al. [14]	Two-tier ensemble PSO based feature selection	93.55				91.67	
Fitriyani et al.[18]	DBSCAN + SMOTE-ENN + XGBoost	0.959	0.971	0.946	0.954	0.953	0.92
Proposed Method	Inf-FS$_s$+BO+IWRF	**0.983**	**0.986**	**0.972**	**0.979**	**0.991**	**0.966**

TABLE 9. Performance evaluation of proposed method compared with previous studies for patient survival using HD clinical record dataset.

Author	Method	Performance Evaluation					
		Accuracy	Precision	Recall	Specificity	F_1	MCC
Chicco [8]	LR	0.838				0.719	0.616
Ishaq et al. [17]	SMOTE + RF + ET	0.926	0.93	0.93	0.93		
Proposed Method	Inf-FS$_s$+BO+IWRF	**0.972**	**0.944**	**0.943**	**0.943**	**0.982**	**0.922**

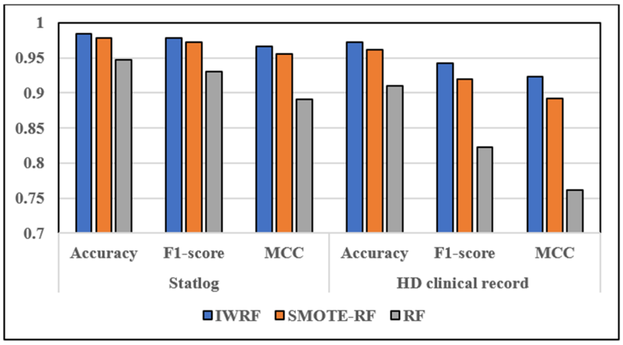


FIGURE 2. The comparison between the proposed IWRF and SMOTE-RF.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{13}$$

$$f = \frac{2 * percision * recall}{percision + recall} \tag{14}$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP+FP)(TP + FN)(TN+FP)(TN + FN)}} \tag{15}$$

A. FEATURE SELECTION RESULTS

Inf-FS$_S$-based feature selection is conducted at each stage of the 10-fold cv utilizing the training data. The Inf-FS$_S$ method ranks and weights each feature in the 13 and 12 features pool for both datasets. Table 2 summarizes the features and associated Inf-FS$_S$ weights for a given fold. The top ten attributes for each validation fold are chosen from these ranking features automatically. Nine characteristics appear in both datasets' top ten features for each of the 10-folds of the training data evaluated. As a result, the presence and survival classifications use these nine features. The selected features for both datasets are listed in Table 3.

B. CLASSIFICATION RESULTS

The developed IWRF model was used for both datasets and showed significant improvement in prediction accuracy compared to existing models. For comparison, we chose six distinct machine learning models (G-NB, LR, SVM, kNN, XGBoost, and RF) frequently utilized in the research field and have a proven record of accuracy and efficiency. The results of different ML models are presented in Table 4 and Table 5 for Statlog and HD clinical records, respectively, including the effects of both with and without FS. IWRF performed better across both datasets than other ML models achieving accuracy, F-measure, and MCC up to 95.5%, 94%, and 0.9 for Statlog, 93.3%, 86%, and 0.81, for the HD clinical dataset, respectively. Also, it is noted that all models have been improved when using FS on both datasets, especially for IWRF, by reaching accuracy, F-measure, and MCC up to 97.7%, 97%, and 0.95 for Statlog, and 95.9%, 91.3%, and 0.88, for HD clinical dataset, respectively.

Furthermore, it can be shown from the results that IWRF achieved better results than the standard RF model in handling the imbalanced data, where IWRF improved the performance for detecting CVD and patients' survival by 3.7% and 5%, respectively, after FS. Recognizing the minority class sufficiently during classification is difficult because the standard RF and the other models used to learn from data input are biased towards the majority class. With the benefit of feature selection, doctors can forecast the survival of patients and the presence of HD by assessing the essential attributes.

To get another point, the IWRF was compared with SMOTE as it is commonly used in handling unbalanced datasets. As with any sampling technique, SMOTE is not a stand-alone classifier but can be integrated with any classifier. For a fair comparison, SMOTE was combined with RF and then compared with IWRF. Table 6 and Table 7 present the results of IWRF against base RF with SMOTE for both datasets. Moreover, we employed BO for tuning SMOTE hyperparameters (sampling ratio and k-neighbors) and (α, p) for IWRF, while the other hyperparameters such n_estimators, max_depth, max_features, and min_samples_split, are set for the default values as in Sklearn library. The findings showed that IWRF achieved higher results than base RF with SMOTE since SMOTE has several drawbacks related to overlap and noisy information. It regularly assigns a global k-neighbor but ignores the local distribution features [48], [49]. The hyperparameter tuning improved model prediction accuracy, but it showed more impact on SMOTE. Increasing the k-neighbor value to compensate for the imbalance ratio may be effective in SMOTE. The results illustrated in Figure 2 show that the improvement achieved by the proposed IWRF is higher than SMOTE-RF compared to the base RF classifier.

The proposed model improved the performance of CVD detection by 3.62%, 4.82%, for the Statlog dataset, and 6.3%, 11.98% for HD clinical records in terms of accuracy and f-measure, respectively.

In the end, we compared our findings to those of previous studies. Because we utilized the same datasets as previous research, we could take the results from prior works without employing their methods. The comparison between the proposed model with the earlier studies is presented in Table 8 and Table 9 for Statlog and HD clinical records datasets. According to this assessment and evaluation, the current research on the CVD detection and survival prediction model outperforms past work, showing an accuracy improvement of 2.4% and 4.6% on Statlog and HD clinical records datasets, respectively. Therefore, the proposed Inf-FSs-BO-IWRF model may be recommended for CVD detection and patients' survival based on the overall findings.

V. CONCLUSION

This article aims to present an accurate and efficient machine learning ensemble model for predicting the presence of CVD and cardiac patient survival. The proposed model integrates Inf-FS$_S$, IWRF, and BO. Those three methods are utilized to select the most significant features, handle the imbalanced data classification issue found in medical datasets, and tune the weighting factor. The developed model is evaluated using two public datasets and benchmarked against previous studies.

The experimental results show that the proposed model was more effective in achieving higher results without changing the data distribution. Also, the proposed IWRF improves the performance of detecting CVD by 3.62% and 6.3% compared to the standard RF.

This research can significantly enhance the healthcare system and serve as a valuable tool for healthcare professionals in diagnosing and forecasting heart failure survival. For our future work, we aim to develop a general framework based on ML ensembles, including outlier detection and removal, and optimize critical hyperparameters of ML ensemble models to improve the detection and severity level classification of various diseases using clinical data.

REFERENCES

[1] R. T. Selvi and I. Muthulakshmi, "An optimal artificial neural network based big data application for heart disease diagnosis and classification model," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 6, pp. 6129–6139, Jun. 2021.

[2] E. J. Benjamin *et al.*, "Heart disease and stroke statistics—2019 update: A report from the American heart association," *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019.

[3] G. Bazoukis, S. Stavrakis, J. Zhou, S. C. Bollepalli, G. Tse, Q. Zhang, J. P. Singh, and A. A. Armoundas, "Machine learning versus conventional clinical methods in guiding management of heart failure patients—A systematic review," *Heart Failure Rev.*, vol. 26, no. 1, pp. 23–34, Jan. 2021.

[4] C. Sowmiya and P. Sumitra, "A hybrid approach for mortality prediction for heart patients using ACO-HKNN," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 5, pp. 1–8, 2020.

[5] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8221–8231, 2015.

[6] K. B. Nahato, K. N. Harichandran, and K. Arputharaj, "Knowledge mining from clinical datasets using rough sets and backpropagation neural network," *Comput. Math. Methods Med.*, vol. 2015, Mar. 2015, Art. no. 460189.

[7] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics Informat.*, vol. 36, pp. 82–93, Mar. 2019.

[8] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–16, Dec. 2020.

[9] B. A. Tama and S. Lim, "A comparative performance evaluation of classification algorithms for clinical decision support systems," *Mathematics*, vol. 8, no. 10, p. 1814, Oct. 2020.

[10] 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 Inf. Syst.*, vol. 2018, pp. 1–21, Dec. 2018.

[11] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, "MIFH: A machine intelligence framework for heart disease diagnosis," *IEEE Access*, vol. 8, pp. 14659–14674, 2019.

[12] M. A. Khan and F. Algarni, "A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS," *IEEE Access*, vol. 8, pp. 122259–122269, 2020.

[13] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019.

[14] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," *BioMed Res. Int.*, vol. 2020, Apr. 2020, Art. no. 9816142.

[15] M. A. Khan, "An IoT framework for heart disease prediction based on MDCNN classifier," *IEEE Access*, vol. 8, pp. 34717–34727, 2020.

[16] M. A. Khan, M. T. Quasim, N. S. Alghamdi, and M. Y. Khan, "A secure framework for authentication and encryption using improved ECC for IoT-based medical sensor data," *IEEE Access*, vol. 8, pp. 52018–52027, 2020.

[17] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021.

[18] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, pp. 133034–133050, 2020.

[19] M. Waqar, H. Dawood, H. Dawood, N. Majeed, A. Banjar, and R. Alharbey, "An efficient SMOTE-based deep learning model for heart attack prediction," *Sci. Program.*, vol. 2021, Mar. 2021, Art. no. 6621622.

[20] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.

[21] M. Bader-El-Den, E. Teitei, and T. Perry, "Biased random forest for dealing with the class imbalance problem," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2163–2172, Jul. 2018.

[22] Z. Jiang, T. Pan, C. Zhang, and J. Yang, "A new oversampling method based on the classification contribution degree," *Symmetry*, vol. 13, no. 2, p. 194, Jan. 2021.

[23] L. Cai, H. Wu, and K. Zhou, "Improved cancer biomarkers identification using network-constrained infinite latent feature selection," *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0246668.

[24] P. Liu and S. Fei, "Two-stage prediction of comorbid cancer patient survivability based on improved infinite feature selection," *IEEE Access*, vol. 8, pp. 169559–169567, 2020.

[25] S. K. Bashar, D. Han, F. Zieneddin, E. Ding, T. P. Fitzgibbons, A. J. Walkey, D. D. Mcmanus, B. Javidi, and K. H. Chon, "Novel density Poincaré plot based machine learning method to detect atrial fibrillation from premature atrial/ventricular contractions," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 448–460, Feb. 2020.

[26] H. Li, Z. Lin, Z. An, S. Zuo, W. Zhu, Z. Zhang, Y. Mu, L. Cao, and J. D. P. García, "Automatic electrocardiogram detection and classification using bidirectional long short-term memory network improved by Bayesian optimization," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103424.

[27] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geosci. Frontiers*, vol. 12, no. 1, pp. 469–477, Jan. 2021.

[28] UCI Machine Learning Repositry. *Statlog (Heart) Data Set*. Accessed: Oct. 21, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/statlog+(heart)

[29] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0181001.

[30] H. Pham and S. Olafsson, "Bagged ensembles with tunable parameters," *Comput. Intell.*, vol. 35, no. 1, pp. 184–203, 2019.

[31] H. Pham and S. Olafsson, "On Cesaro averages for weighted trees in the random forest," *J. Classification*, vol. 37, no. 1, pp. 223–236, 2020.

[32] C. Chen, A. Liaw, and L. Breiman, "Using random forest to learn imbalanced data," *Univ. California, Berkeley*, vol. 110, nos. 1–12, p. 24, 2004.

[33] S. Xuan, G. Liu, and Z. Li, "Refined weighted random forest and its application to credit card fraud detection," in *Proc. Int. Conf. Comput. Social Netw.*, Shanghai, China, Nov. 2018, pp. 343–355.

[34] V. Y. Kulkarni and P. K. Sinha, "Effective learning and classification using random forest algorithm," *Int. J. Eng. Innov. Technol.*, vol. 3, no. 11, May 2014.

[35] L. I. Kuncheva and J. J. Rodriguez, "A weighted voting framework for classifiers ensembles," *Knowl. Inf. Syst.*, vol. 38, no. 2, pp. 259–275, Feb. 2014.

[36] K. Gajowniczek, I. Grzegorczyk, T. Ząbkowski, and C. Bajaj, "Weighted random forests to improve arrhythmia classification," *Electronics*, vol. 9, no. 1, p. 99, Jan. 2020.

[37] M. S. Babu and V. Vijayalakshmi, "An effective approach for sub-acute ischemic stroke lesion segmentation by adopting meta-heuristics feature selection technique along with hybrid naive Bayes and sample-weighted random forest classification," *Sens. Imag.*, vol. 20, no. 1, pp. 1–24, Dec. 2019.

[38] L. V. Utkin, A. V. Konstantinov, V. S. Chukanov, M. V. Kots, M. A. Ryabinin, and A. A. Meldo, "A weighted random survival forest," *Knowl. Based Syst.*, vol. 177, pp. 136–144, Aug. 2019.

[39] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," *Neural Comput. Appl.*, vol. 29, no. 10, pp. 685–693, 2018.

[40] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli, "Infinite latent feature selection: A probabilistic latent graph-based ranking approach," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1398–1406.

[41] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, and M. Cristani, "Infinite feature selection: A graph-based feature filtering approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4396–4410, Dec. 2020.

[42] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," 2012, *arXiv:1206.2944*.

[43] E. Hazan, A. Klivans, and Y. Yuan, "Hyperparameter optimization: A spectral approach," 2017, *arXiv:1706.00764*.

[44] N. DeCastro-García, L. M. Castañeda, D. E. García, and M. V. Carriegos, "Effect of the sampling of a dataset in the hyperparameter optimization phase over the efficiency of a machine learning algorithm," *Complexity*, vol. 2019, Feb. 2019, Art. no. 6278908.

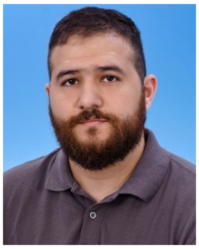
[45] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 24, 2011, pp. 1–9.

[46] R. Elshawi, M. Maher, and S. Sakr, "Automated machine learning: State-of-the-art and open challenges," 2019, *arXiv:1906.02287*.

[47] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Montreal, QC, Canada, 1995, vol. 14, no. 2, pp. 1137–1145.

[48] J. A. Sáez, J. Luengo, J. Stefanowski, and F. Herrera, "SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," *Inf. Sci.*, vol. 291, pp. 184–203, Jan. 2015.

[49] K. Cheng, C. Zhang, H. Yu, X. Yang, H. Zou, and S. Gao, "Grouped SMOTE with noise filtering mechanism for classifying imbalanced data," *IEEE Access*, vol. 7, pp. 170668–170681, 2019.



ABDALLAH ABDELLATIF (Member, IEEE) received the M.Sc. degree in control systems from Universiti Tun Hussein Onn Malaysia, in 2019. He is currently pursuing the Ph.D. degree in control systems with the Department of Electrical Engineering, Universiti Malaya. His research interests include data mining, mainly working machine learning and deep learning.

HAMDAN ABDELLATEF (Member, IEEE) received the B.E. and M.Sc. degrees in communication and electronics engineering from Beirut Arab University, Lebanon, in 2012 and 2016, respectively, and the Ph.D. degree in electrical engineering from Universiti Teknologi Malaysia, Malaysia, in 2020. He is currently a Research Faculty and a Postdoctoral Research Fellow with Lebanese American University, Byblos, Lebanon. His current research interests include deep learning, hardware/software co-design, signal processing, and stochastic computing.

JEEVAN KANESAN received the degree in electrical engineering from UTM, in 1999, and the M.Sc. and Ph.D. degrees from USM, Penang, in 2002 and 2006, respectively. He joined the Department of Electrical Engineering, Universiti Malaya, in 2008. After obtaining his degree, he worked as an Engineer at Carsem Semiconductor, Ipoh, before resuming his master's degree. After obtaining his Ph.D. degree, he worked as a Research and Development Engineer at Intel, Penang, for two and half years. Currently, he has published more than 70 peer reviewed journals. His research interests include optimization and machine learning.

CHEE-ONN CHOW (Senior Member, IEEE) received the Bachelor of Engineering (Hons.) and Master of Engineering Science degrees from the University of Malaya, Malaysia, in 1999 and 2001, respectively, and the Doctor of Engineering degree from Tokai University, Japan, in 2008. He joined the Department of Electrical Engineering, University of Malaya, in 1999, where he is currently an Associate Professor. His research interests include communication networks, multimedia applications, data analytics, and artificial intelligence. He is a Registered Professional Engineer with the Board of Engineers Malaysia.

JOON HUANG CHUAH (Senior Member, IEEE) received the B.Eng. (Hons.) degree from Universiti Teknologi Malaysia, the M.Eng. degree from the National University of Singapore, and the M.Phil. and Ph.D. degrees from the University of Cambridge. He is currently the Head of the VIP Research Group and an Associate Professor with the Department of Electrical Engineering, Faculty of Engineering, University of Malaya. He is a Chartered Engineer registered under the Engineering Council, U.K., and also a Professional Engineer registered under the Board of Engineers, Malaysia. He was the Honorary Treasurer of IEEE Computational Intelligence Society (CIS) Malaysia Chapter and the Honorary Secretary of IEEE Council on RFID Malaysia Chapter. He is the Vice-Chairperson of the Institution of Engineering and Technology (IET) Malaysia Network. He is also a fellow and the Honorary Secretary of the Institution of Engineers, Malaysia (IEM). His research interests include image processing, computational intelligence, IC design, and scanning electron microscopy.

HASSAN MUWAFAQ GHENI received the Bachelor (B.Sc.) degree in electrical and electronic engineering from the Department of Electrical Engineering, Babylon University, Hilla, Iraq, in June 2016. In February 2018, he entered the Master's Program with the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Malaysia. He is currently a Lecturer with the Department of Computer Techniques Engineering, Al-Mustaqbal University College. His research interests include optical communication, the IoT, wireless sensor networks, communications, V2V systems, and artificial intelligence.