Shahrood University of Technology

Research paper

Application of Stacked Ensemble Techniques in Head and Neck Squamous Cell Carcinoma Prognostic Feature Subsets

D. K. OWUSU[1*], Ch. C. Nyarko[2], J. Acquah[2] and J. Yarney[3]

1. University of Mines and Technology, Ghana.
2. Department of Mathematical Sciences, Faculty of Engineering, University of Mines and Technology, Tarkwa, Ghana.
3. Department of Radiotherapy and Oncology, National Centre for Radiotherapy & Nuclear Medicine, Korle Bu, Accra, Ghana.

Article Info	Abstract
	Head and neck cancer (HNC) recurrence is ever increasing among Ghanaian men and women. Because not all machine learning classifiers are equally created, even if multiple of them suite very well for a given task, it may be very difficult to find one which performs optimally given different distributions. The stacking learns how to best combine weak classifier models to form a strong model. As a prognostic model for classifying HNSCC recurrence patterns, this study tried to identify the best stacked ensemble classifier model when the same ML classifiers for feature selection and stacked ensemble learning are used. Four stacked ensemble models; in which first one uses two base classifiers: gradient boosting machine (GBM) and distributed random forest (DRF); second one uses three base classifiers: GBM, DRF, and deep neural network (DNN); third one uses four base classifiers: GBM, DRF, DNN, and generalized linear model (GLM); and fourth one uses five base classifiers: GBM, DRF, DNN, GLM, and Naïve bayes (NB) were developed, using GBM meta-classifier in each case. The results show that implementing stacked ensemble technique consisting of five base classifiers on gradient boosted features achieves better performance than achieves on other feature subsets, and implementing this stacked ensemble technique on gradient boosted features achieves better performance compared to other stacked ensemble techniques implemented on gradient boosted features and other feature subsets used. Learning stacked ensemble technique having five base classifiers on GBM features is clinically appropriate as a prognostic model for classifying and predicting HNSCC patients' recurrence data.

1. Introduction

Treatment of recurrent Head and Neck Cancer (HNC) requires accurate prognosis associated with it in order to determine the type and extent of therapy for effective and to fully destroy cancerous cells in the human body. The recurrence rate of HNSCC is ever increasing, which decreases the survival rate, and yet there are few studies on the applications of Machine Learning (ML) techniques in the prognosis of recurrent Head and Neck Squamous Cell Carcinoma (HNSCC). Previous studies have shown that there is an increase in survival rates of patients with HNSCC as a result of advances in treatments as well as modification of lifestyle [9,21]. The increase in survival rates might be associated with secondary aftermaths: recurrent HNSCC (local recurrence, regional recurrence, and distant recurrence) or second primary tumor [10,46]. In spite of the significant improvement in treatment modalities to increase overall survival (OS) of patients with HNSCC, they still experience relapse which affects their survival rates [6,22,43]. The 5-year OS rate was 83% and 48% for patients with HNSCC relapse in the primary stage and advanced stage respectively

[20]. With treatment options like chemotherapy, radiotherapy, surgery, or a combination of these for patients with HNSCC, 40%-50% of them experience recurrences; and presently, the majority of HNCs diagnosed are regionally advanced with lymph node metastases [21,42]. Even with the advances in the treatment modalities of HNSCCs, nearly 650,000 patients with new cases of the disease are diagnosed each year globally, and one third of these patients experience relapse or recurrence [2,4,34,45]. A recently published global cancer statistics also reported that each year, new HNSCC cases of more than 800,000 are diagnosed [6]. Also, during the follow-up tests, 10% to 20% of patients with early-stage cancer experience recurrent HNSCC, while a recurrence rate of approximately 50% was experienced by those with locally advanced stage of the disease, particularly in locoregional pattern [1]. These HNCs have annual incidence worldwide of more than 550,000 cases accounting for 300,000 deaths each year [22]. It is the sixth most common cancer worldwide, and more than half a million of new cases emerge [17,43]. Approximately 375,000 number of deaths from HNC worldwide was recorded in 2012, of which 4.6% was total cancer mortality [17]. The major cause of morbidity is recurrent HNSCC, and is reducing long-term survival of patients with HNSCC. HNC–related death is mostly contributed by locoregional recurrences, which is 15% - 50% of patients with HNSCC [5,7,34]. Over the decades, various ML techniques have been applied in the cancer diagnosis and prognosis. Medically, the disease is identified in a patient by its signs and symptoms (called diagnosis) and the prediction of its outcome on the patient is studied (called prognosis). Different subtypes of cancer have been identified and categorized as a heterogeneous disease. The application of ML techniques has been the aim to define a model for the progression and treatment of cancer subtypes. Various standalone ML techniques, including but not limited to Artificial Neural Network (ANNs), Naïve Bayes (NB), Support Vector Machines (SVMs), and Decision Trees (DTs); and homogeneous ensemble ML techniques including Gradient Boosting Machine (GBM) and Random Forest (RF), have been applied in a wide range of various cancer research to build prognostic models from complex datasets, known to offer effective and high accuracy in decision making; thus, revealing their importance [26]. Based on the latest PubMed statistics, over 1,500 published articles on cancer with ML techniques application have been recorded. Nonetheless, majority of these papers focused on the applications of ML techniques to

identify, classify, detect, and/or distinguish tumors and other malignancies. Primarily, ML techniques have been applied to cancer diagnosis and detection [32]. According to [12,26], the applications of ML techniques for cancer prediction and prognosis have only been relatively recently used by cancer researchers. As a result, the number of published papers in this field is relatively less studied; thus, less than 120 papers published. Studies by [3,13-15,19,25,29,30,35] had proved that ML techniques are powerful to generate more accurate diagnosis or prognosis outcomes that conventional statistical methods could not otherwise generate. Given less published papers on application of ML techniques in HNSCC prognosis, most researchers focused on HNSCC susceptibility and/or survivability, with very few of them focusing on HNSCC recurrence. Examples of published articles on ML applications in recurrent HNSCC subtypes are; [3,8,11,16,23,31,37,39-41,48].

[3] applied a feed-forward ANN, where LR was the benchmark, to identify the prognosis of locoregional recurrences in early-stage oral tongue squamous cell carcinoma (OTSCC), and concluded that ANN with accuracy of 92.7% outperformed LR with accuracy of 86.5%. [8] identified the prognostic factors for locally advanced nasopharyngeal carcinoma relapse, and concluded that DT classifiers showed high prediction in the prognosis of individual recurrence pattern with overall accuracy of 84.5% – 95.2% compared to the Kaplan-Meier analysis. [40,41,48] applied SVM in the prediction of laryngeal cancer recurrence and concluded that the said prognostic model had high accuracy, and thus a good classifier for tumor progression. [37] applied SVM in the prognosis of oral cancer recurrence. [36] used SVM with accuracy of 87.0% to identify the prognosis of patients with recurrence and metastasis of HNSCC and concluded to be a good model for HNSCC recurrence/metastasis. A study on nasopharyngeal carcinoma for local recurrence was conducted by [11], with 7 classification techniques; DT, KNN, LDA, LR, NB, RF and RBF-SVM, and 6 feature selection techniques; MIM, FSCR, RELF-F, CMIM, MRMR, and JMI. They achieved optimal integration methods of outperforming classifiers for prognosis with accuracies of (FSCR+RF: 89.2%, FSCR+KNN: 88.3%, FSCR+RBF-SVM: 86.7%, and MRMR+RBF-SVM: 88.3%). A similar study conducted by [31] identified the prognosis of patients with nasopharyngeal carcinoma who were treated with intensity-modulated radiotherapy (IMRT) and experienced recurrence, using ANN, KNN, and SVM, and concluded that the

classification models considered exhibited potential and high prediction accuracies; ANN: 81.2%, KNN: 77.5%, and SVM: 73.2%. Again, [16] identified the most prognostic features that are associated with oral squamous cell carcinoma (OSCC) relapses using Dynamic Bayesian Network (DBN), NB, ANN, SVM, DT, and RF with accuracies obtained on clinical data as (BN: 73.7%, NB: 74.6%, ANN: 74.6%, SVM: 74.6%, DT: 81.6%, RF: 74.6%), on imaging data (BN: 86.4%, NB: 87.5%, ANN: 83%, SVM: 84.1%, DT: 77.3%, RF: 83%), on tissue genomic data (BN: 75.8%, NB: 74.2%, ANN: 74.2%, SVM: 74.2%, DT: 69.2%, RF: 80%), and on blood genomic data (BN: 87.5%, NB: 91.7%, ANN: 95.8%, SVM: 95.8%, DT: 87.5%, RF: 87.5%). They concluded that, with all the three-input (clinical, imaging, and genomic) data fed into DBN, the model revealed a high accuracy of 100%.

Though these previous studies produced some useful results, each using one or several ML techniques on individually basis for recurrent HNSCC subtypes prognosis, it is possible that their prognostic models would have been outperformed by an ensemble ML technique if they had been combined to the training data in a stacking ensemble; given now that, there is a rapid significant development of ensemble ML techniques in cancer prediction and prognosis. According to [12], not all ML techniques are equally created; while some perform better given problems of certain kinds, others do better given problems of other kinds. For instance, given the size of biological domain, some ML techniques may scale nicely to meet such size, while others will not. Similarly, some techniques might have assumptions regarding the kind of data that might render them incapable for a given problem at hand. It is this that makes it more important to consider stacked generalization of ML techniques on any given set of training data in cancer prognosis.

Many well-known modern ML techniques are mostly ensembles; including bagging (random forest) and boosting (gradient boosting machine) and have been applied in most studies, particularly in medical fields to obtain better performance [18,23,38]. Whereas random forest minimizes the variance and solves over-fitting issues in the model, boosting minimizes the bias and solves under-fitting issues in the model during the training process. Stacking combines multiple of different base classifiers into a strong one in their combination using a meta learning algorithm. The benefit of stacking is that it can harness the capabilities of a range of well-performing models on a classification task and make predictions that have better performance than any single model in the ensemble [47]. It worthwhile to combined these techniques in an ensemble learning so as to achieve a well optimally ensemble classification model for HNSCC recurrence among HNSCC patients using stacked ensemble learning of ML classifiers.

2. Materials and Methods
2.1. Dataset
To evaluate the performance of the classification models, the HNSCC subtypes dataset including laryngeal cancer, hypopharyngeal cancer, nasopharyngeal cancer, and oropharyngeal cancer was obtained from the registry of radiotherapy and oncology department at Korle Bu, Accra. It has a total of 125 instances, 18 attributes (features), and a class label with binary outcome coded 1 (as recurrence) or 0 (as nonrecurrence). There are 33 and 92 female and male records respectively. The data was preprocessed using model imputation to avoid the deletion of instances with missing examples. In order to normalize the training examples, one-hot encoding was used on attributes with more than two levels. The summary of this dataset is shown table 1. To generate stacked ensemble model, [23] proposed a stacked ensemble algorithm, a technique that found the optimal weighted average of diverse base learners for classification of various healthcare datasets (Wisconsin Breast Cancer, Pima Indian Diabetes Dataset, and Indian Liver Patient Dataset using GBM, DRF, and DNN as base learners, and GLM as a meta learner to stack GBM and DRF in one case; then stack GBM, DRF, and DNN in another case. [27] likewise proposed a stacked ensemble algorithm, a technique that found the best meta-learner in a stacking ensemble for classifying breast cancer, using GBM, DRF, DNN, and GLM as base learners, and each of which was used as a meta-learner to determine the best meta-learner in a stacking ensemble. Base on the ML algorithms considered by [23,27] as the most effective algorithms to providing the most effective ensemble classification model for HNSCC prognosis, all have been employed under this study with the inclusion of NB to experiment a stacked ensemble consisting of five (5), at least one more that of the *state-of-the-art* stacked ensemble model consisting of a maximum of four (4) base classifiers in HNC prognosis. Thus, NB was chosen from among the most effective single base classifiers (DT, KNN, NB, and SVM) considered by the previous studies, based on its performance on the experimental data. Data augmentation was generally used to improve a model's performance. It is a technique that comprises a set of methods

used to artificially increase the number of data samples present in the dataset. This was done as deep learning models generalize well when the number of data samples available to train on is large. In this way, state-of-the-art models can be created with fewer data samples available. The data augmentation technique is usually applied to computer vision applications where domain-specific data, such as medical data, is not abundantly available. Thus, data augmentation technique was used. In medical research, it takes time to collect sufficient samples as most patients are usually lost to follow-up to check whether or not they had a recurrence and thus, the sample size is usually small. HNSCC is considered recurrence if the patient was treated with curative intent and after the cancer reaches its remission, they redeveloped HNSCC termed as recurrence. Patients that received palliative treatment intent and still had cancer are not considered cancer recurrent patients. Unfortunately, most patients received palliative intent treatment and only a few could receive curative intent due to financial difficulties, causing small instances. Hence, there is a need to implement feature selection methods to identify significant variables that are important to the clinical outcomes and to avoid the over-fitting problem. In this research, implementing the feature selection method aims to find an optimal number of features for the small sample of oral cancer prognosis data. The number of features in the dataset was considered too many (18 attributes) if compared to the sample size (125 instances). Thus, the feature selection method is needed to reduce the number of features and select only those significant to HNSCC prognosis. Thus, the original dataset was subjected to five feature selection techniques, namely GBM, DRF, DNN, GLM, and NB, each provided feature subset of the data as shown table

2. Training data (75%) and test data (25%) were constructed for each data subset. The evaluation metrics; accuracy, recall, specificity, logarithmic loss (logloss), and AUC of Receiver Operating Characteristic Curve were used to measure the performance of the classification models. The H2o package for machine learning library in R programming language was used.

3. Description of Proposed Stacked Ensemble Techniques

This paper presented four different techniques of stacked ensemble learning. The first one used two base classifiers, namely gradient boosting machine (GBM) and distributed random forest (DRF); the second one used three base classifiers, namely GBM, DRF, and deep neural network (DNN); the third one used four base classifiers, namely GBM, DRF, DNN, and generalized linear model (GLM); and the fourth one used five base classifiers, namely GBM, DRF, DNN, GLM, and Naïve bayes (NB); and in each case, a meta-classifier called GBM was used [27].

Various cancer data subsets related to HNSCC provided by various feature selection techniques used in this study were used, and compare the performance of stacked ensemble models on these various data subsets.

The evaluation results confirmed that stacked ensemble techniques built on Gradient Boosted feature subset (GBM-FS) has the ability to perform better compared to stacked ensemble techniques built on feature subsets provided by other feature selection techniques. Similarly, the evaluation results confirmed that stacked ensemble techniques consisting of five base classifiers has the ability to perform better compared to other stacked ensemble techniques considered on five feature subsets of HNSCC dataset.

Table 1. Dataset description.

Dataset	No. of instances	No. of attributes	Class label with No. of instances
HNSCC	125	18	Class 1: recurrence (61); class 0: nonrecurrence (60)

Table 2. Optimal Feature Subset Selected.

Feature Selection Technique	Feature Subset Selected
GBM-FS	Nodes, Age, Smoke, StageIV, p63, TreatCCRT, PaTT4, Size
DRF-FS	TreatCCRT, Age, Smoke, Invasion, PlNN2, HPV, PaTT2, TreatRT, Nodes, PaTT4, StageIV, SiteNPC, p16, Size, Drink,
DNN-FS	TreatCCRT, TreatRT, p63, Nodes, p16, Size, Smoke, HPV
GLM-FS	p63, TreatCCRT, StageII, paTT3, Smoke, StageIII, Nodes, PlNN3
NB-FS	TreatCCRT, p63, Smoke, Nodes, paTT3, TreatRT, Invasion, Age
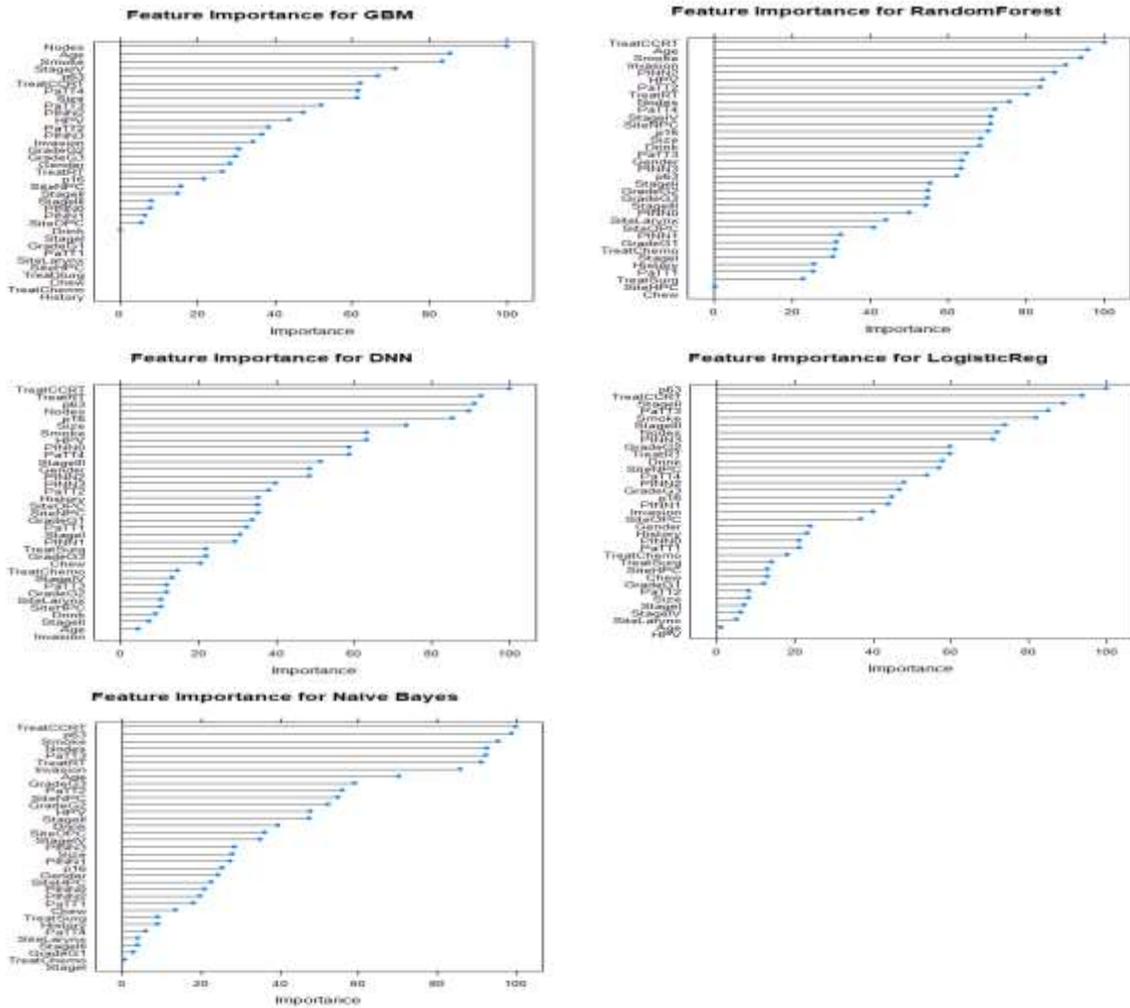
Figure 1. Rank plots for features by feature selection techniques.

Algorithm 1. Stacking with k-fold (k=10) cross validation

Input: Dataset $D = \{\mathbf{x}_i, y_i\}_{i=1}^n$; learning rate $\alpha > 0$
$C = \{h_1, h_2, \ldots, h_L\}$ – classifiers set which constitute the ensemble.

Output: An ensemble classifier H

Step 1: Adopt cross validation approach in preparing a training set for meta-classifier
Randomly split D_s into V equal-size subsets: $D = \{D_1, D_2, \ldots, D_K\}$
for $v \leftarrow 1$ to K **do**
    Step 1.1: Learn first-level classifiers $\{h_1, h_2, \ldots, h_L\}$
    for $l \leftarrow 1$ to L **do**
        Learn a classifier h_{kl} from D/D_k
    end for
    Step 1.2: Construct a training set for second-level classifiers
    for $\mathbf{x}_i \in D_k$ **do**
        Get a record $\{\mathbf{x}'_i, y_i\}$, where $\mathbf{x}'_i = \{h_{k1}(\mathbf{x}_i), h_{k2}(\mathbf{x}_i), \ldots, h_{kL}(\mathbf{x}_i)\}$
    end for
end for
Step 2: Learn second-level classifier
    Re-learn first-level classifier h'_l from the collection of $Z = \{\mathbf{x}'_i, y_i\}_{i=1}^n$

end for
Return $H(\mathbf{x}) = h'\left(h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_L(\mathbf{x})\right)$
Step 3: Predict unseen example (testing set)

for each $\mathbf{x} \in D_t$ **do**
    Apply an ensemble classifier $H(\mathbf{x})$ on $\mathbf{x}$.
end for

To achieve better performance using these base classifiers from H2o, GBM, DRF, DNN, GLM, and NB were selected [44]. For the meta-classifier, GBM model was used [27] as it was the best performing base classifier among the base classifiers considered in this study as shown in Figure 2. To obtain data subsets for learning stacked ensemble techniques, each base classifier was used to perform feature selection, each of

which ranked the features according to their importance; and using 60% threshold, feature subsets were obtained as shown in Figure 1 and table 2 respectively. The algorithm 1 shows the learning of stacked ensemble models with 10-fold cross-validation based on the hyperparameters obtained by random grid search as shown in Table 3.
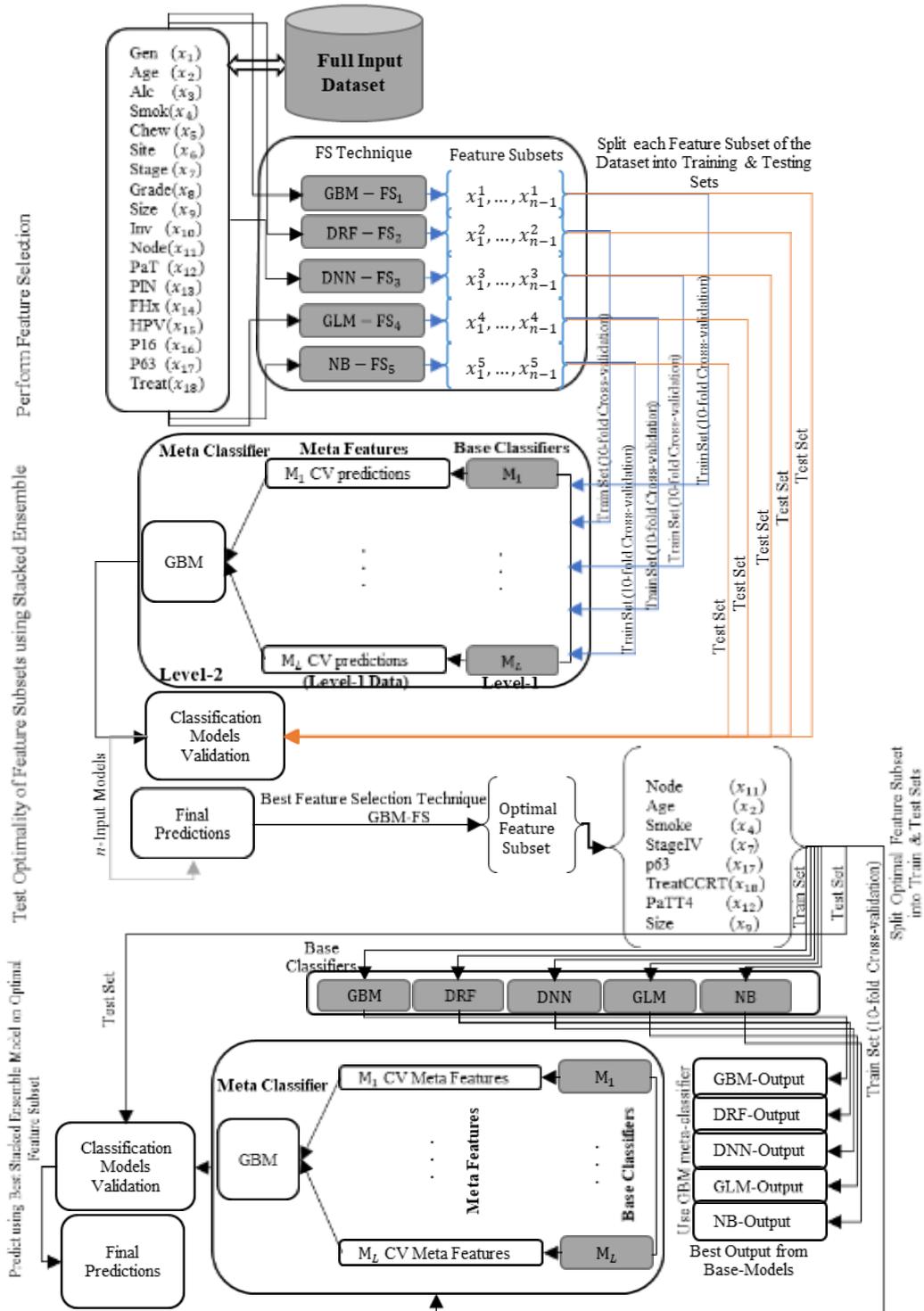
Figure 2. Flowchart of a proposed stacked ensemble technique.

Table 3. Classifiers with their corresponding hyper-parameter values.

Classifiers	Hyper-parameters in grid search with the corresponding range of values	hyperparameters fixed values
GBM	max_depth = c(7, 9), learn_rate = c(0.01, 0.1), learn_rate_annealing=c(0.99, 1), sample_rate=c(0.5, 0.7, 1), col_sample_rate=c(0.8, 0.9, 1)	ntrees = 3000 nfolds = 10 Fold_assignment = "Modulo" keep_cross_validation_predictions = True stopping_rounds = 50
DRF	max_depth = c(9, 30), mtries = 3, sample_rate = c(0.5, 0.75, 1), col_sample_rate_per_tree= (0.8, 0.9, 1)	ntrees = 3000 nfolds = 10 Fold_assignment = "Modulo" keep_cross_validation_predictions = True stopping_rounds = 50
DNN	activation=c("Rectifier", "Tanh"), hidden = c(5, 10, 50), l1 = c(0, 1e-3, 1e-5), l2 = c(0, 1e-3, 1e-5),	epochs = 20 nfolds = 10 Fold_assignment = "Modulo" keep_cross_validation_predictions = True stopping_rounds = 50
NB	Laplace=c(0, 5, by 0.5)	nfolds = 10 Fold_assignment = "Modulo" keep_cross_validation_predictions = True
GLM	alpha=c(0.1)	nfolds = 10 remove_collinear_columns = True Fold_assignment = "Modulo" keep_cross_validation_predictions = True

4. Results

This study compared the performance of the stacked ensemble techniques implemented on various feature subsets of the HNSCC dataset provided by various feature selection techniques used in this study. The stacked ensemble techniques were trained on the training set, and were evaluated on the test set for each data subset.

Table 4. Performance of stacked ensemble model (model-GBM2) consisting of two base classifiers (GBM AND DRF) on test data.

Metrics	Feature selectors				
	GBM-FS	DRF-FS	DNN-FS	GLM-FS	NB-FS
Accuracy	0.8172	0.7813	0.6667	0.7419	0.6875
Logloss	0.3379	0.5879	0.7379	0.5959	0.7889
Recall	0.8939	0.8636	0.9130	0.9245	0.8421
Specificity	0.6296	0.6000	0.4255	0.5000	0.4615
AUC	0.8018	0.7391	0.7289	0.7684	0.7198

Table 4 shows the performance of the proposed the stacked ensemble technique having two base classifiers (GBM and DRF) on the test set of different feature subsets of the data; table 5 shows the performance of the proposed stacked ensemble technique having three base classifiers (GBM, DRF, and DNN) on the test set for all the data subsets used in this study; Table 6 shows the performance of the proposed stacked ensemble technique having four base classifiers (GBM, DRF, DNN, and GLM) on the test set for all the data subsets used in this study; and Table 7 shows the performance of the proposed stacked ensemble technique having five base classifiers (GBM, DRF,

DNN, GLM, and NB) on the test set for all the data subsets used in this study.

Table 5. Performance of stacked ensemble model (model-GBM3) consisting of three base classifiers (GBM, DRF, AND DNN) on test data.

Metrics	Feature selectors				
	GBM-FS	DRF-FS	DNN-FS	GLM-FS	NB-FS
Accuracy	0.8278	0.7813	0.7527	0.7312	0.7813
Logloss	0.3267	0.4167	0.4257	0.4267	0.4269
Recall	0.9206	0.8333	1.0000	0.8793	0.8333
Specificity	0.6333	0.6250	0.9855	0.4857	0.6250
AUC	0.8625	0.8623	0.7023	0.7322	0.7319

Table 6. Performance of stacked ensemble model (model-GBM4) consisting of four base classifiers (GBM, DRF, DNN, and GLM) on test data.

Metrics	Feature selectors				
	GBM-FS	DRF-FS	DNN-FS	GLM-FS	NB-FS
Accuracy	0.8817	0.8438	0.8280	0.7742	0.7500
Logloss	0.3042	0.4141	0.4111	0.4241	0.4441
Recall	0.9143	0.6667	0.8630	0.9138	0.8571
Specificity	0.7826	0.9500	0.7000	0.5429	0.5455
AUC	0.8809	0.8179	0.7597	0.7908	0.7536

Considering the Tables 4, 5, 6, and 7, for the data subsets used in this study, best results were obtained using stacked ensemble learning. For the stacked ensemble having two base classifiers on test data in table 4, best accuracy (81.72%), log loss (0.3379), specificity (62.96%), and AUC (0.8018) were obtained for data subset provided by GBM feature selection technique. For stacked ensemble model having three base classifiers on test data in

table 5, best accuracy (82.78%), log loss (0.3267), and AUC (0.8625) for data subset provided by GBM feature selection technique. Interestingly, best recall (100%) and specificity (98.55%) were obtained for DNN data subset.

Table 7. Performance of stacked ensemble model (model-GBM5) consisting of five base classifiers (GBM, DRF, DNN, GLM, AND NB) on test data.

	Feature selectors				
Metrics	GBM-FS	DRF-FS	DNN-FS	GLM-FS	NB-FS
Accuracy	0.9063	0.8817	0.8172	0.8280	0.8438
Logloss	0.2959	0.3041	0.3679	0.3659	0.3359
Recall	0.7500	0.9265	0.8714	0.8533	0.9091
Specificity	1.0000	0.7600	0.6522	0.7222	0.7000
AUC	0.9251	0.8321	0.7947	0.8010	0.8333

For stacked ensemble model having four base classifiers on test data in Table 6, best accuracy (88.17%), log loss (0.3042), recall (91.43%), and AUC (0.8809) was obtained for GBM data subset. Best specificity (95.00%) was obtained for DRF data subset. For stacked ensemble model having five base classifiers on test data in Table 7, best accuracy (90.63%), log loss (0.2959), specificity (100%), and AUC (0.9251) was obtained for GBM data subset. Best recall (92.65%) was obtained for DRF feature subset. The graphs of the information in tables 4, 5, 6, and 7 are represented in Figure 3.
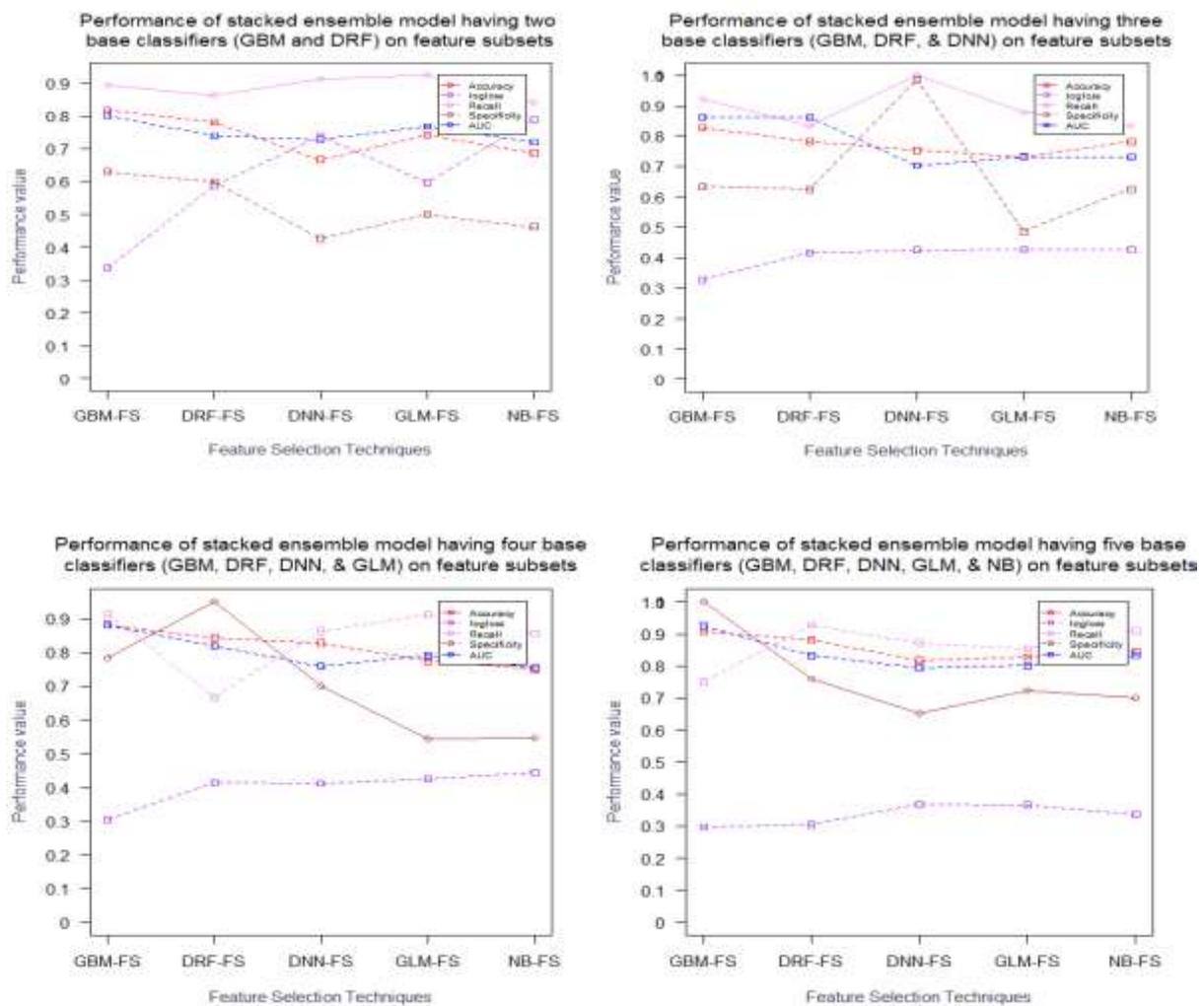
Figure 3. Performance plots of feature selection techniques based on stacked ensemble techniques.

In addition, Tables 8, 9, 10, 11, and 12 show the performance comparison of various stacked ensemble techniques implemented on each feature subset of the data used in this study. For data subsets provided by each feature selection technique, best results were obtained using stacked ensemble learning. Table 8 shows the performance comparison of various stacked ensemble techniques implemented on the test set of features subset provided by GBM feature selection.

Table 8. Performance comparison of stacked ensemble models on GBM-FS test set.

Metrics	Stacked ensemble model-GBM2	Stacked ensemble model-GBM3	Stacked ensemble model-GBM4	Stacked ensemble model-GBM5
Accuracy	0.8172	0.8278	0.8817	0.9063
Logloss	0.3379	0.3267	0.3042	0.2959
Recall	0.8939	0.9206	0.9143	0.7500
Specificity	0.6296	0.6333	0.7826	1.0000
AUC	0.8018	0.8625	0.8809	0.9251

Table 9. Performance comparison of stacked ensemble models on DRF-FS test set.

Metrics	Stacked ensemble model-GBM2	Stacked ensemble model-GBM3	Stacked ensemble model-GBM4	Stacked ensemble model-GBM5
Accuracy	0.7813	0.7813	0.8438	0.8817
Logloss	0.5879	0.4167	0.4141	0.3041
Recall	0.8636	0.8333	0.6667	0.9265
Specificity	0.6000	0.6250	0.9500	0.7600
AUC	0.7391	0.8623	0.8179	0.8321

Table 10. Performance comparison of stacked ensemble models on DNN-FS test set.

Metrics	Stacked ensemble model-GBM2	Stacked ensemble model-GBM3	Stacked ensemble model-GBM4	Stacked ensemble model-GBM5
Accuracy	0.6667	0.7527	0.8280	0.8172
Logloss	0.7379	0.4257	0.4111	0.3679
Recall	0.9130	1.0000	0.8630	0.8714
Specificity	0.4255	0.9855	0.7000	0.6522
AUC	0.7289	0.7023	0.7597	0.7947

Table 11. Performance comparison of stacked ensemble models on GLM-FS test set.

Metrics	Stacked ensemble model-GBM2	Stacked ensemble model-GBM3	Stacked ensemble model-GBM4	Stacked ensemble model-GBM5
Accuracy	0.7419	0.7312	0.7742	0.8280
Logloss	0.5959	0.4267	0.4241	0.3659
Recall	0.9245	0.8793	0.9138	0.8533
Specificity	0.5000	0.4857	0.5429	0.7222
AUC	0.7684	0.7322	0.7908	0.8010

Table 12. Performance comparison of stacked ensemble models on NB-FS test set.

Metrics	Stacked ensemble model-GBM2	Stacked ensemble model-GBM3	Stacked ensemble model-GBM4	Stacked ensemble model-GBM5
Accuracy	0.6875	0.7813	0.7500	0.8438
Logloss	0.7889	0.4269	0.4441	0.3359
Recall	0.8421	0.8333	0.8571	0.9091
Specificity	0.4615	0.6250	0.5455	0.7000
AUC	0.7198	0.7319	0.7536	0.8333

It can be observed that the stacked ensemble technique having five base classifiers performed better than other techniques implemented on GBM feature subset of the data used in this study. For this data subset, best accuracy (90.63%), log-loss (0.2959), specificity (100%), and AUC (0.9251) were obtained using stacked ensemble technique having five base classifiers followed by stacked ensemble technique having four base classifiers with accuracy (88.17%), logloss (0.3042), and AUC (0.8809). Best recall (92.06%) was obtained using stacked ensemble technique having three base classifiers. In table 9, best accuracy (88.17%),

log loss (0.3041), recall (92.65%), and AUC (0.8321) were obtained using the stacked ensemble technique having five base classifiers followed by stacked ensemble technique having four base classifiers with accuracy (84.38%) and log loss (0.4141) for DRF feature subset of the data. For DNN feature subset data, best accuracy (82.80%) with higher log loss (0.4111) were obtained using stacked ensemble technique consisting of four base classifiers followed by stacked ensemble technique having five base classifiers; accuracy (81.72%) with best log loss (0.3679) and AUC (0.7947). For GLM data subset, best accuracy (82.80%), log loss

(0.3659), specificity (72.22%), and AUC (0.8010) were obtained using stacked ensemble technique having five base classifiers followed by stacked ensemble technique having four base classifiers with accuracy (77.43%). For NB data subset, best accuracy (84.38%), log loss (0.3359), recall (90.91%), specificity (70.00%), and AUC (0.8333) were obtained using stacked ensemble technique

having five base classifiers followed by stacked ensemble technique having three base classifiers with accuracy (78.13%), log loss (0.4269), and specificity (62.50%). The graphs of the information in Tables 8, 9, 10, 11, and 12 are represented in Figure 4.
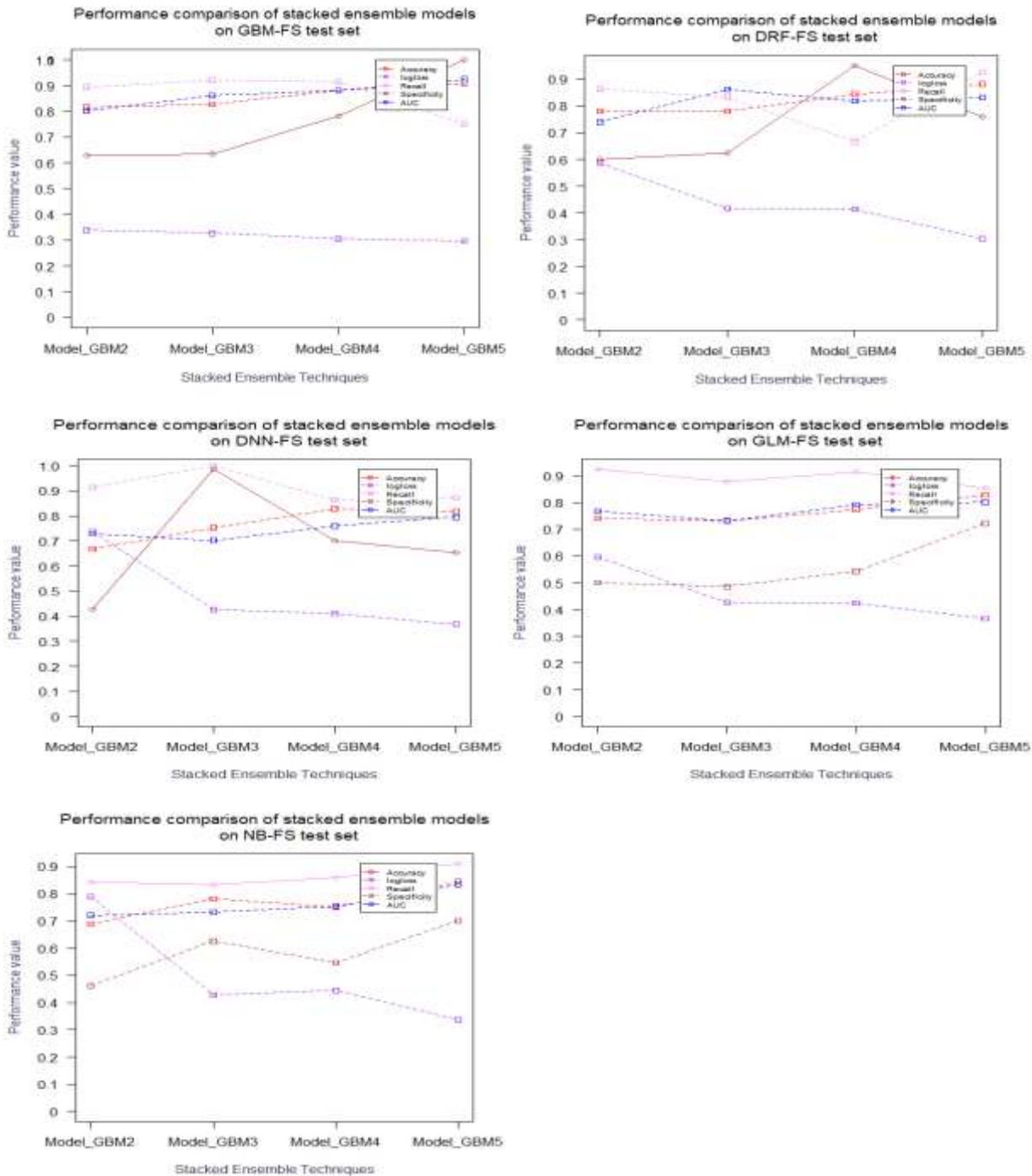
Figure 4. Performance plots of stacked ensemble techniques on various feature subsets.

Partial Dependence Plot (PDP) and Individual Conditional Expectations (ICE) were used to make

prognostications of recurrence patterns on HNSCC patients. The PDP is similar to ICE, and shows the

marginal effect a feature has on the predicted class label (binary classification in this case) of a machine learning model. A PDP or ICE can show whether the relationship that exists between the target and a feature is linear, monotonic or more complex. The yellow curves in figure 5 indicate the PDP while the black curves represent the ICE. The PDP shows how the average prediction of all instances are associated with a feature while the ICE shows how the prediction of each instance is associated with the feature.
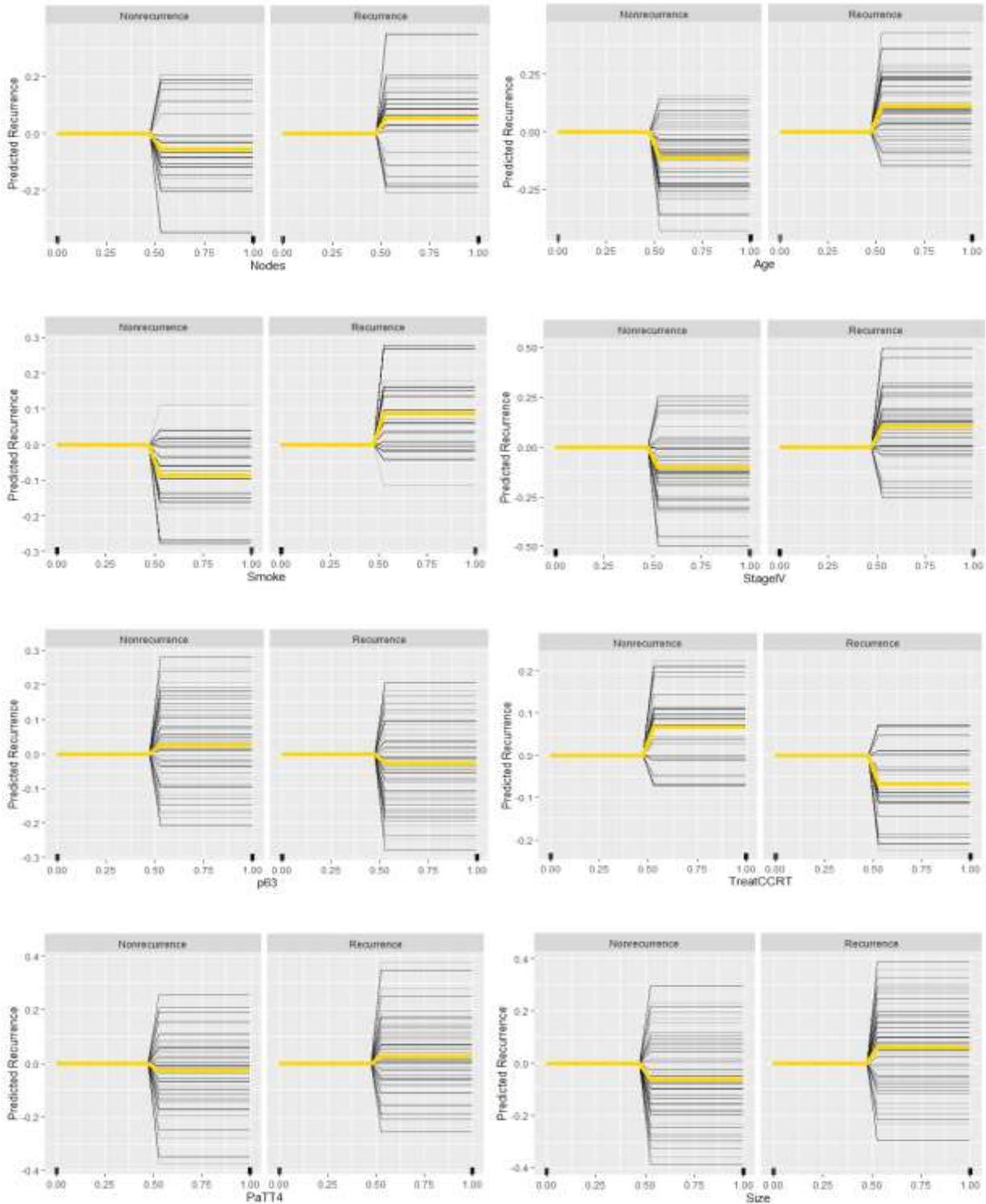
Figure 5. PDP and ICE on features.

For feature *Nodes* on the class label, PDP shows that, the probability of recurrence increases with a greater number of neck nodes exceeding about 50% compared to that with a smaller number of neck nodes. The prediction of recurrence is centred at "0" until the number of lymph nodes exceeds 50%. The probability of recurrent increases around 50% of the presence of lymph nodes, but is this true for every patient (instance) in the dataset? The ICE plot reveals that for most patients, the lymph nodes effect follows the average pattern of an increase of lymph nodes around 50%, but there are some exceptions: For some patients that have a high predicted probability at few of lymph nodes, the predicted recurrence does not change with much presence of lymph nodes.

Similarly, for the feature *Size* on the class label, the PDP explains that with low tumor size of 2 cm or less, the possibility of experiencing recurrence is zero while it is around 0.09 for larger tumor size greater than 2 cm. Interestingly, the predicted probability of experiencing recurrence does not fall when the size of tumor is greater than 2 cm. PDP shows that the recurrent probability increases at around 0.55 (around 2 cm) of the tumor size, but is this true for every instance in the dataset? The ICE plot reveals that for most patients the tumor size effect follows the average pattern of an increase at around 2 cm, but there are some exceptions: For some patients that have a high predicted probability at a smaller tumor size (2 cm), the predicted recurrent HNSCC probability does not change with size (tumor size 2 cm or greater). This feature has a positive marginal effect on the target binary class. For feature *TreatCCRT* on class label, PDP explains that as more of Concurrent Chemoradiotherapy (CCRT) treatment is administered on patients on regular basis, less recurrences are experienced and vice versa. The PDP explains that the probability of recurrence decreases around when the treatment process is half-way (55%) to its completion, but is this true for every patient in the dataset? The ICE plot reveals that for most patients, the TreatCCRT effect follows the average pattern of decrease at around 55% to its completion, but there are some exceptions: For some patients that have low predicted probability at the half-way of treatment with CCRT, the predicted recurrence does not change with TreatCCRT completion. This feature has a positive marginal effect on the target binary class. Similar interpretation can be made for other PDP and ICE plots.

In summary, the results of various stacked ensemble techniques implemented on feature subsets of the data provided by various feature selection techniques used in this study showed that, all the stacked ensemble techniques used in this study achieved higher performance on data subset provided by GBM feature selection technique compared to their performance results on data subsets provided by other feature selection techniques. It was also observed that for each stacked ensemble technique implemented on each feature subset of the data provided by each feature selection technique with the exception of DNN feature subset data, the stacked ensemble technique consisting of five base classifiers achieved the highest accuracy compared to other stacked ensemble techniques used in this study. The stacked ensemble technique having four base classifiers achieved the highest accuracy on data subset provided by DNN feature selection technique. In terms of log loss, it was observed that for each stacked ensemble technique implemented on each feature subset of the data provided by each feature selection technique, the stacked ensemble technique consisting of five base classifiers achieved the least logloss compared to other stacked ensemble techniques used in this study. In terms of AUC, it was also observed that for each stacked ensemble technique implemented on each feature subset of the data provided by each feature selection technique with the exception of DRF feature subset data, the stacked ensemble technique consisting of five base classifiers achieves the highest AUC compared to other stacked ensemble techniques used in this study. The stacked ensemble technique having three base classifiers achieves the highest AUC on data subset provided by DRF feature selection technique. The stacked ensemble model having five base classifiers implemented on GBM feature subset data was used to make predictions using PDP and ICE.

5. Conclusion and Future Work
This paper focused on the improvement of the ensemble classification performance through stacked generalization towards the prediction of HNSCC recurrence patterns using data subsets provided by various feature selection techniques considered in this study. To achieve this, the stacked ensemble technique that finds the optimal weighted average of diverse machine learning base models using meta learning algorithm was used. For base classifiers, GBM and DRF were used and another base classifier DNN along with the previous two (GBM and DRF) was integrated. Next, another base classifier GLM along with the previous three (GBM, DRF, and DNN) was integrated. Then, another base classifier NB along with the previous four (GBM, DRF, DNN, and

GLM) was integrated. To achieve the optimal combination of these diverse of base classifiers models considered in this study, the GBM was used as meta-classifier based on its high performance when trained as base classifier on various data subsets compared to other base classifiers considered in this study. The experimental results showed that using stacked ensemble technique having five base classifiers had better performance compared to other stacked ensemble techniques considered in this study, and using GBM feature selection technique is better as a supporting tool for generating the most accurate prognostic features for HNSCC dataset.

In our future study, we will extend each stacked ensemble learning technique to a multi-level stacked ensemble learning, with multiple layers of stacking at each layer by considering more than a maximum of five base classifiers.

Acknowledgement

Data and materials availability

Contact the corresponding author by email on reasonable request as the dataset generated and analysed for this study was obtained from medical records of KBTH and not publicly available.

Ethical approval

Ethical clearance with the refence number KBTH/MD/G3/21 was obtained from the KBTH Institutional Review Board (KBTH-IRB) after the Scientific and Technical Committee (KBTH-STC) approval with identification number KBTH-STC 000114/2020 for the collection of retrospective data.

References

[1] A. Argiris, K. J. Harrington, M. Tahara, et al. (2017). Evidence-based treatment options in recurrent and/or metastatic squamous cell carcinoma of the head and neck. *Front Oncol.*, vol. 7, no. 72, [PMC free article] [PubMed] [Google Scholar]

[2] A. Argiris, M. V. Karamouzis, D. Raben, and R. L. Ferris, (2008). Head and neck cancer. *Lancet*, vol. 371, no. 9625, pp. 1695–1709.

[3] R. O.Alabi, M. Elmusrati, I. Sawazaki-Colone, L. P. Kowalski, C. Haglund, and Coletta, et al. (2019). Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a Web-based prognostic tool, *VirchowsArchiv*.

[4] J. A. Bonner, P. M. Harari, J. Giralt, R. B. Cohen, C. U. Jones, R. K. Sur, et al. (2010). Radiotherapy plus Cetuximab for Locoregionally Advanced Head and Neck Cancer: 5-year Survival Data from a phase 3 Randomised Trial, and Relation between Cetuximab-Induced Rash and Survival. *The Lancet Oncology*. vol. 11, no. 1, pp. 21–28.

[5] J. Bourhis, A. Le Maître, B. Baujat, H. Audry, and J. P. Pignon. (2007). Meta-Analysis of Chemotherapy in Head, Neck Cancer Collaborative Group, Meta-Analysis of Radiotherapy in Carcinoma of Head, Neck Collaborative Group, Meta-Analysis of Chemotherapy in Nasopharynx Carcinoma Collaborative Group Individual patients' data meta-analyses in head and neck cancer. *Curr. Opin. Oncol.*, vol. 19, pp. 188–194.

[6] F. Bray, et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, vol. 6, no. 68, pp. 394.

[7] B. Brockstein, D. J. Haraf, A. W. Rademaker, M. S. Kies, K. M Stenson, F. Rosen, et al. (2004). Patterns of failure, prognostic factors and survival in locoregionally advanced head and neck cancer treated with concomitant chemoradiotherapy: a 9-year, 337-patient, multi-institutional experience. *Annals of oncology*, vol.15, pp. 1179–1186.

[8] H. Cai, X. Pang, D. Dong, Y. Ma, Y. Huang, X. Fan, et al. (2019). Molecular Decision Tree Algorithms Predict Individual Recurrence Pattern for Locally *Advanced Nasopharyngeal Carcinoma. Journal Cancer*, vol. 10, no. 15, pp. 3323-3332.

[9] E. E. Cohen, S. J. LaMonte, N. L. Erb, K. L.Beckman, N. Sadeghi, , K. A. Hutcheson, et al. (2016). American Cancer Society Head and Neck Cancer Survivorship Care Guideline. *CA Cancer Journal of Clin.*, vol. 66, no. 3, pp. 203 – 239.

[10] T. R. Cooney, and M. G. Poulsen, (1999). Is routine follow-up useful after combinedmodality therapy for advanced head and neck cancer? *Arch. Otolaryngol Head Neck Surg.*, vol. 125, no. 4, pp. 379 – 382.

[11] D. Du, et al. (2019). Machine Learning Methods for Optimal Radiomics-Based Differentiation. *jnl of Molecular Imaging & Biology·*.

[12] J. A. Cruz, and D. S. Wishart, (2006). Review: Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*, vol. 2, pp. 59-74.

[13] A. Diamant, A. Chatterjee, M. Vallières, G. Shenouda, and J. Seuntjens, (2019). Deep learning in head & neck cancer outcome prediction, *scientific reports*, pp. 1-19 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6391436/.

[14] R. M. Dom, S. A. Kareem, R. Zain and B. Abidin, (2007). An Adaptive Fuzzy Regression Model for the Prediction of Dichotomous Response Variables. *Paper presented at the Fifth International IEEE Conference on Computational Science and Applications*.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, (2001). Pattern classification (2nd edition). *New York: Wiley.*

[16] K. P. Exarchos, Y. Goletsis, and D. I. Fotiadis, (2012). Multiparametric decision support system for the prediction of oral cancer reoccurrence. *IEEE Trans InfTechnol Biomed*, vo. 16, no. 6, pp. 1127–34.

[17] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, et al. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer,* vol. 136, no. 5, pp. 359–386.

[18] C. Gibbons, et al. (2017). Supervised machine learning algorithms can classify open-text feedback of doctor performance with human level accuracy. *J. Med. Internet Res*. vo. 19, no. 3.

[19] P. Giraud, P. Giraud, A. Gasnier, R. El Ayachy, S. Kreps, J. P. Foy, C. Durdux, F. Huguet, A. Burgun, and J. E. Bibault, (2019). Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers. *Frontiers. Oncology*. vol. 9, no. 174, pp. 9:174.

[20] W. J. Goodwin, (2000). Salvage surgery for patients with recurrent squamous cell carcinoma of the upper aerodigestive tract: when do the ends justify the means? *Laryngoscope*, vol. 110, no. 3 (Pt 2 Suppl 93), pp. 1–18.

[21] R. I. Haddad, and D. M. Shin, (2008). Recent advances in head and neck cancer. Natl. Engl. *Journal for Med.*, vol. 359, pp. 1143–1154.

[22] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, (2011). Global cancer statistics. *CA Cancer J. Clin.*, vol. 61, no. 2, pp. 69–90.

[23] F. Kabir, and S. A. Ludwig, (2019), "Enhancing the Performance of Classification Using Super Learning", *Data-Enabled Discov*. Appl., vol. 3, no. 5, pp. 2-13.

[24] H. Kaur, S. Batra, (2017). HPCC an Ensembled framework for the prediction of the onset of diabetes. *International Conference on Signal Processing, Computing and Control.*

[25] T. Kawazu, A. Kazuyuki, K. Yoshiura, E. Nakayama and S. Kanda, (2003). Application of Neural Networks to the Prediction of Lymph Node Metastasis in Oral Cancer. *Oral Radiol*, vol. 19, pp. 137-142.

[26] K. Kourou, T. P. Exarchosa, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, (2015). Review: Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8-17.

[27] H. Kwon, J. Park, and Y. Lee, (2019). Stacking Ensemble Technique for Classifying Breast Cancer. *Jnl of Health Informatics Research*, vol. 25, no. 4, pp. 283-288.

[28] L. Lavanya, and J. Chandra, (2019). Oral Cancer Analysis Using Machine Learning Techniques. *International Journal of Engineering Research and Technology*, vol. 12, no. 5, pp. 596-601.

[29] D. C. Li, H. C. Hsu, T. I. Tsai, T. J. Lu, and S. Hu, (2007a). A New Method to Help Diagnose Cancers for Small Sample Size. *Expert Systems with Applications*, vol. 33, pp. 420- 424.

[30] H. Li, D. Li, C. Zhang, and S. Nie, (2007b). An Application of Machine Learning in the Criterion Updating of Diagnosis Cancer. *International Conference in Neural Networks and Brain 2005*, vol. 1, pp. 187-190.

[31] S. Li, K. Wang, Z. Hou, J. Yang, W. Ren, S. Gao, et al. (2018). Use of Radiomics Combined with Machine Learning Method in the Recurrence Patterns After Intensity-Modulated Radiotherapy for Nasopharyngeal Carcinoma: A Preliminary Study. *Journal Frontiers in Oncology.*

[32] J. F. McCarthy, K. A. Marx, P. E. Hoffman, et al. (2004). Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. *Ann N Y AcadSci*, vol. 1020, pp. 239-262.

[33] T. M. Mitchell, (2006). The discipline of machine learning. Carnegie Mellon University, School of Computer Science, Machine Learning Department.

[34] M. R. Posner, D. M. Hershock, C. R. Blajman, E. Mickiewicz, E. Winquist, V. Gorbounova, et al. (2007). Cisplatin and fluorouracil alone or with docetaxel in head and neck cancer. *Natl. Engl. Journal of Med.*, vol. 357, no. 17, pp. 1705–1715.

[35] V. S. H. Rao, and M. N. Kumar, (2011). A New Intelligence Based Approach for Computer-Aided Diagnosis of Dengue Fever. *IEEE Transactions on Information Technology in Biomedicine*, (In press).

[36] I. P. Ribeiro, F. Caramelo, L. Esteves, J. Menoita, F. Marques, L. Barroso, et al. (2017). Genomic predictive model for recurrence and metastasis development in head and neck squamous cell carcinoma patients. *Scientific Report*, pp. 1-8.

[37] P. Rosado, P. Lequerica-Fernández, L. Villallaín, I. Peña, F. Sanchez-Lasheras, and J. C. de Vicente, (2013). Survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers and support vector machines. *Expert SystAppl*, vol. 40, pp. 4770–4776.

[38] T. Silwattananusarn, W. Kanarkard, K. Tuamsuk, (2016). Enhanced classification accuracy for cardiotocogram data with ensemble feature selection and classifier ensemble. *J. Comput. Commun.* vol. 4, bo. 4, pp. 20.

[39] A. Singh, S. Goyal, Y. J. Rao, and M. Loew, (2019). Tumor Heterogeneity and Genomics to Predict Radiation Therapy Outcome for Head-and-Neck Cancer: A Machine Learning Approach. *Radiation Oncology*, vol. 105, no. 1, pp. 232-233.

[40] J. Su, Y. Zhang, H. Su, C. Zang, and W. Li, (2017). A recurrence model for laryngeal cancer based on SVM and gene function clustering. *Acta Otolaryngol*, vol. 137, no. 5, pp. 557-562.

[41] Z. Tang, G. Wei, L. Zhang, and Z. Xu, (2019). Signature microRNAs and long noncoding RNAs in laryngeal cancer recurrence identified using a competing endogenous RNA netwoek. Molecular Medicine Reports, *Spandidos publications*, pp. 4806-4818.

[42] G. Tolstonog, and C. Simon, (2017). Trends in surgical research in head and neck cancer. Curr. Treat. *Options Oncol.*, vol. 18, pp. 38.

[43] L. A Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, (2015). Global cancer statistics, 2012. *CA Cancer Journal Clin.*, vol. 65, no. 2, pp. 87–108.

[44] J. Vanerio, and P. Casas, (2017). Ensemble-learning approaches for network security and anomaly detection. Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, *ACM*.

[45] J. B. Vermorken, E. Remenar, C. van Herpen, T. Gorlia, R. Mesia, M. Degardin, et al. (2007). Cisplatin, fluorouracil, and docetaxel in unresectable head and neck cancer. *Natl. Engl. Journal Med.*, vol. 357, no. 17, pp. 1695–1704.

[46] K. S. Viro, R. C. Paniello, and F. E. Johnson, (1998). Costs of posttreatment surveillance for patients with upper aerodigestive tract cancer. *Arch. Otolaryngol Head Neck Surg.*, vol. 124, no. 5, pp. 564 – 572.

[47] L. Yalang, G. Jing, L. QI, and F. Wei, (2015). Ensemble Learning. pp. 484-504.

[48] B. Yag, Q. Guo, F. Wang, K. Cai, X. Bao, and J. A. Chu, (2017). 80-gene set potentially predicts the relapse in laryngeal carcinoma optimized by support vector machine. *Cancer Biomarkers*, vol. 19, no. 1, pp. 65-73.

[49] Yarey et al. (2017). Cancers of the Head & Neck: Does cncurrent chemoradiotherapy preceded by chemothrapy improve survival in locally advanced nasopharyngeal cancer patients? Experience from Ghana. *BioMed Central*, vol. 4, no. 2, pp. 1-7.

Owusu و همکاران

مجله هوش مصنوعی و داده‌کاوی، دوره دوازدهم، شماره اول، سال ۱۴۰۳.

کاربرد تکنیک‌های مجموعه‌ای انباشته در زیرمجموعه‌های ویژگی پیش آگهی کارسینوم سلول سنگفرشی سر و گردن

Damianus Kofi Owusu[*.1]، **Christiana Cynthia Nyarko**[2]، **Joseph Acquah**[2] و **Joel Yarney**[3]

۱ دانشگاه معدن و صنعت، غنا.

۲ گروه علوم ریاضی، دانشکده فنی، دانشگاه معادن و فناوری، تارکوا، غنا.

۳ بخش رادیوتراپی و انکولوژی، مرکز ملی رادیوتراپی و پزشکی هسته ای، کورله بو، آکرا، غنا..

چکیده:

عود سرطان سر و گردن (HNC) در میان مردان و زنان غنا در حال افزایش است .از آنجایی که همه طبقه‌بندی‌کننده‌های یادگیری ماشین به طور یکسان ایجاد نمی‌شوند، حتی اگر چندین طبقه‌بندی از آنها برای یک کار معین بسیار مناسب باشند، ممکن است یافتن طبقه‌بندی‌کننده‌ای که توزیع‌های مختلف را به‌طور بهینه انجام می‌دهد بسیار دشوار باشد .انباشته یاد می گیرد که چگونه مدل های طبقه‌بندی کننده ضعیف را به بهترین شکل ترکیب کند تا یک مدل قوی را تشکیل دهد .به عنوان یک مدل پیش‌آگهی برای طبقه‌بندی الگوهای عود HNSCC ، این مطالعه سعی کرده است بهترین مدل طبقه‌بندی گروه پشته‌ای را زمانی که از طبقه‌بندی‌کننده‌های ML مشابه برای انتخاب ویژگی و یادگیری گروه انباشته استفاده می‌شود، شناسایی کند. چهار مدل مجموعه انباشته ؛که در آن اولی از دو طبقه‌بندی‌کننده پایه استفاده می‌کند: ماشین تقویت گرادیان (GBM) و جنگل تصادفی توزیع شده (DRF). دومی از سه طبقه‌بندی کننده پایه استفاده می‌کند GBM :، DRF، و شبکه عصبی عمیق (DNN). سومی از چهار طبقه‌بندی کننده پایه استفاده می‌کند GBM :، DRF، DNN، و مدل خطی تعمیم یافته (GLM). و چهارم از پنج طبقه‌بندی‌کننده پایه استفاده می‌کند GBM ، DRF، DNN، GLM، و بیزین ساده (NB) با استفاده از متا طبقه‌بندی‌کننده GBM . در هر مورد نتایج نشان می‌دهد که پیاده‌سازی تکنیک مجموعه انباشته متشکل از پنج طبقه‌بندی کننده پایه بر روی ویژگی‌های تقویت‌شده گرادیان، عملکرد بهتری نسبت به سایر زیرمجموعه‌های ویژگی به دست می‌آورد، و اجرای این تکنیک مجموعه‌ای انباشته بر روی ویژگی‌های تقویت‌شده گرادیان، عملکرد بهتری را در مقایسه با سایر تکنیک‌های گروه پشته‌ای اجرا شده بر روی شیب تقویت‌شده به دست می‌آورد. تکنیک یادگیری مجموعه انباشته دارای پنج طبقه‌بندی کننده پایه بر روی ویژگی‌های GBM از نظر بالینی به عنوان یک مدل پیش آگهی برای طبقه‌بندی و پیش‌بینی داده‌های عود بیماران HNSCC مناسب است.

کلمات کلیدی: تکرار HNSCC ، یادگیری ماشینی، پشته‌بندی، انتخاب ویژگی، طبقه‌بندی.