

Article

Diagnosis and Prediction of Large-For-Gestational-Age Fetus Using the Stacked Generalization Method

Faheem Akhtar ^{1,2} , Jianqiang Li ^{1,*} , Yan Pei ³ , Azhar Imran ¹ , Asif Rajput ² ,
Muhammad Azeem ¹  and Qing Wang ⁴

¹ School of Software Engineering, Beijing University of Technology, Beijing Engineering Research Center for IoT Software and Systems, Beijing 100124, China; fahim.akhtar@iba-suk.edu.pk (F.A.); azharimran63@gmail.com (A.I.); mazeem.qau@hotmail.com (M.A.)

² Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan; asifali@iba-suk.edu.pk

³ School of Computer Science and Engineering, the University of Aizu, Aizuwakamatsu 965-8580, Japan; peiyan@u-aizu.ac.jp

⁴ Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China; qing.wang@tsinghua.edu.cn

* Correspondence: lijianqiang@bjut.edu.cn; Tel.: +86-106-7396-842

Received: 30 August 2019; Accepted: 10 October 2019; Published: 14 October 2019



Abstract: An accurate and efficient Large-for-Gestational-Age (LGA) classification system is developed to classify a fetus as LGA or non-LGA, which has the potential to assist paediatricians and experts in establishing a state-of-the-art LGA prognosis process. The performance of the proposed scheme is validated by using LGA dataset collected from the National Pre-Pregnancy and Examination Program of China (2010–2013). A master feature vector is created to establish primarily data pre-processing, which includes a features' discretization process and the entertainment of missing values and data imbalance issues. A principal feature vector is formed using GridSearch-based Recursive Feature Elimination with Cross-Validation (RFECV) + Information Gain (IG) feature selection scheme followed by stacking to select, rank, and extract significant features from the LGA dataset. Based on the proposed scheme, different features subset are identified and provided to four different machine learning (ML) classifiers. The proposed GridSearch-based RFECV+IG feature selection scheme with stacking using SVM (linear kernel) best suits the said classification process followed by SVM (RBF kernel) and LR classifiers. The Decision Tree (DT) classifier is not suggested because of its low performance. The highest prediction precision, recall, accuracy, Area Under the Curve (AUC), specificity, and F1 scores of 0.92, 0.87, 0.92, 0.95, 0.95, and 0.89 are achieved with SVM (linear kernel) classifier using top ten principal features subset, which is, in fact higher than the baselines methods. Moreover, almost every classification scheme best performed with ten principal feature subsets. Therefore, the proposed scheme has the potential to establish an efficient LGA prognosis process using gestational parameters, which can assist paediatricians and experts to improve the health of a newborn using computer aided-diagnostic system.

Keywords: recursive feature elimination with cross-validation; stacked generalization; large for gestational age; feature engineering; machine learning; bioinformatics

1. Introduction

During the last several decades, an increased in the incidence of LGA neonates in the developed countries are reported, which is even wider in developing countries [1,2]. It is defined as a fetus whose gestational weight is above the 90th percentile of a fetus with a similar gestational age and sex [3]. It exhibits serious pre/post maternal complications, which are comprised of shoulder dystocia [4,5],

insulin resistance [6,7], metabolic syndrome [6], prolonged labor [5], cesarean section [8], postpartum bleeding [8], serious adverse consequences before and after delivery including breast cancer [9,10], and expedited chances of infants mortality rate [11]. Therefore, based on these serious complications, which are associated with the health of a newborn, it is a topic of keen interest for paediatricians and related health-care officials.

On the basis of the above-discussed concerns, the primary motivation behind this research is to develop an accurate LGA classification model, which is capable of classifying an LGA fetus before birth using maternal biochemical indicators. To improve LGA classification performance, using the national pre-pregnancy and examination program of China (2010–2013) dataset [12], the master feature vector (MFV) is created to formalize LGA dataset, where we discretized feature values and entertained missing values. Principal feature subsets were created using the proposed GridSearch-based RFECV+ IG feature selection scheme followed by stacking to select, extract, and rank features to enhance proposed classification schemes performance with minimal generalization error. Based on the experimental results, it is foreseen that the top ten features selected by each of the feature selection processes proved best, but Support Vector Machine (SVM) with linear kernel remained best with the production of highest performance metrics scores. Moreover, to establish comparative analysis, the proposed scheme is compared with previously published research on the same LGA dataset.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 defines the methodology of this research with complete details of data pre-processing, experimental flow, and performance metrics. Section 4 presents experimental results of various experimental processes. Section 5 discusses experimental results and compares them with existing baselines schemes to signify the importance of the proposed scheme. Finally, this paper is concluded in Section 6 and future work is presented.

2. Related Work

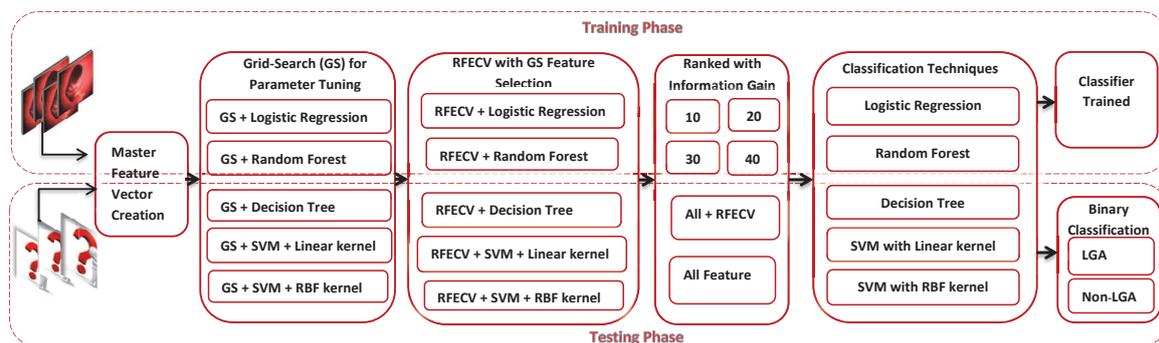
Previously, different practices were used to identify LGA fetus. The most common methods were using estimated fetal weight (EFW), abdominal circumferences (AC), ultrasound surveillance of an obese woman, maternal BMI, gestational weight, gestational diabetes mellitus (GDM), etc. Most of them were observational or retrospective studies that used simple logistic regression to extract discriminant features subset for the establishment of an LGA prognosis process. For example, Shen used sonographic estimated fetal weight (EFW) of Chinese women to classify a fetus as LGA or non-LGA and achieved specificity and sensitivity of 48.1% and 97.3% respectively [13]. Blue used AC with EFW for LGA classification [14]. Harper proposed to use ultrasound surveillance of obese women before 32 weeks of the gestational period to classify a fetus as LGA or non-LGA [15]. Chen used maternal BMI with gestational weight for LGA classification [16]. Moore established a cohort analysis and demonstrated that LGA fetus exhibits dichotomous risks at term [17]. Luangkwan used linear modelling to observe the risk of parental complications in pregnant women with an LGA fetus [18]. In addition, some research was proposed to monitor variations in fetus bio-chemical indicators during different physical checkups to control its consequences [19–21]. Based on this, it can easily be foreseen that most of them were observational or retrospective studies that used simple logistic regression to extract discriminant features subset for the establishment of an LGA prognosis process.

Perhaps, in our previous work, we were the first who exploited machine learning (ML) techniques for the establishment of an efficient LGA prognosis process. In [22], we used information gain (IG) feature selection scheme for the LGA prognosis and achieved precision and Area Under the Curve (AUC) score of 0.71 and 0.70, respectively. In [23], we used IG with an ensemble scheme to improve classification performance with the extraction of useful features to establish an efficient LGA prognosis process where we achieved precision and AUC scores of 0.84 and 0.72 respectively. Furthermore, in [24], using experts' expertise, we reached to obtain prediction precision and AUC scores of 0.95 and 0.86, respectively. In this research, we identified ranked twenty in practice features for the establishment of an efficient LGA prognosis process. However, still, there is still room to improve

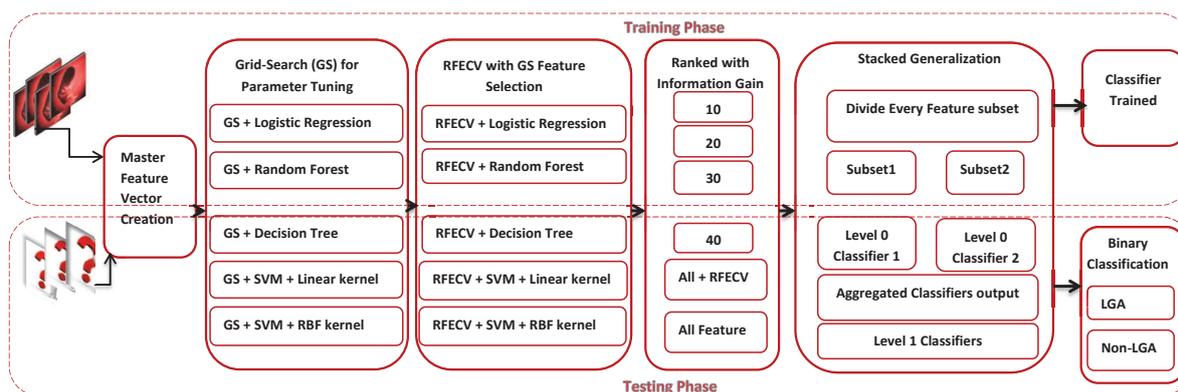
the prediction performance of an LGA classification system. Therefore, a master feature vector is created and GridSearch-based Recursive Feature Elimination with Cross-validation (RFECV) scheme followed by stacked generalization is introduced to select, rank, extract suitable feature subset with higher classification prediction performance with reduced generalization errors. RFECV and stacked generalization have previously been proven best in various related application domains [25,26].

3. Materials and Methods

This research proposes two different schemes for LGA classification. In the first scheme, creation of a master feature vector, GridSearch-based Recursive feature elimination with cross-validation (RFECV) feature selection scheme with machine learning models that are tuned with GridSearch, and ranked features subset based on Information gain (IG) feature selection scheme are given to four influential machine learning classifiers, as shown in Scheme 1, which illustrates the methodology of the first proposed LGA classification scheme. The second scheme is intended to enhance the LGA classification performance with minimized generalization errors. The objective is to expedite LGA classification performance with an ensemble of stacked classifiers based on the meta-level features extracted from the level-0 of the stacking procedure. Later, the extracted features from Level-0 of stacking process is given to Level-1 of the stacking process to establish a state of the art LGA classification model. Refer to Scheme 2, which illustrates the methodology of the proposed LGA classification scheme. In both of the schemes, the classifiers are constructed and tested with ten-fold cross-validation to diagnose an infant as LGA or non-LGA. The reason to deploy ten-fold cross-validation is to minimize the generalization errors and come up with a standardized LGA classification framework.



Scheme 1. The complete methodology of the proposed GridSearch + Recursive Feature Elimination with Cross-validation + Information Gain-based feature selection scheme for the establishment of an efficient Large for Gestational Age infants prognosis process.



Scheme 2. The complete methodology of the proposed GridSearch+Recursive Feature Elimination with Cross-validation + Information Gain + Information Gain + Stacked generalization based feature selection and classification scheme for the establishment of an efficient Large for Gestational Age infants prognosis process with reduced generalization error.

3.1. Dataset Collection

The benchmark LGA dataset used in this research is collected from National Pre-Pregnancy and Examination Program of China [12]. The program was initiated to eliminate birth deficiencies of Chinese citizens across China (2010 to 2013). The project covered all of the provisional and municipal hospitals of China. The examination checklist was suggested and finalized by the mutual consensus of a panel of experts constituted from various related domains (i.e., obstetrics, paediatrics, andrology, internal medicine, etc.). The checklist includes pre-pregnancy items (i.e., eating habits (male(m)/female(f)), smoking (m/f), drinking (m/f), height (m/f), occupation (m/f), etc.), pregnancy items that includes parents’ (clinical measures, reproductive system measures, abnormalities in pregnancy, etc.) and socio-economic and demographic factors.

The obtained dataset is comprised of 371 features with 215,568 records. The distribution of the data is accomplished with a widely used LGA classification scheme proposed by Zhu et al. [27]. It is presented in Table 1. Based on the proposed scheme, each of the records is classified either as LGA or non-LGA. Therefore, following the proposed scheme, 26,226 records are labelled with the LGA, and remaining 189,342 are labelled with the non-LGA.

Table 1. Birthwise Large for Gestational Age infants classification chart which is a widely used and accepted guideline on the Chinese population.

Birth Week	Boys Weight (g)	Girls Weight (g)	Birth Week	Boys Weight (g)	Girls Weight (g)
24	846	740	34	2843	2768
25	1031	939	35	3114	3028
26	1212	1132	36	3386	3286
27	1390	1321	37	3637	3515
28	1566	1504	38	3828	3691
29	1742	1686	39	3979	3803
30	1925	1872	40	4030	3872
31	2122	2071	41	4092	3921
32	2341	2285	42	4148	3963
33	2584	2519	-	-	-

3.2. Preparation of the Master Feature Vector

For improving LGA classifiers performance, the master feature vector in the LGA dataset is required to be accurate, robust, and flawlessly identified. As mentioned previously, the LGA dataset is obtained from an official project that was launched across almost every related hospital in China; and it is evident that every massive project always contains a certain amount of missing fields. The reason

might not always be a human error, but certain times, a paediatrician did not feel to prescribe or record specific test that is mentioned in the proposed recording guidelines. The entertainment of these missing fields is a significant and necessary step during the classification task. Otherwise, it will adversely affect the classification results. Therefore, considering the desired need for the establishment of a better LGA classification model, the following algorithm is proposed to eliminate discussed issues.

Above defined Algorithm 1 can be explained as follows, Let's suppose L_0 is the basic LGA dataset with f_0 features and n_0 records. To create MFV, initially, a classification column is added to each record as per [27] classification criteria. Each feature value is discretized with the help of literature and paediatrician's expertise; a threshold to delete 10% from controls (LGA's record) and 15% from cases (non-LGA) records are set and remaining records were imputed with the selected mode of these feature values. As a result of this process, a Master Feature Vector (MFV) is extracted from the complete LGA dataset. The details of the resultant MFV is presented in Figure 1. where, Figure 1a, represents the distribution of original LGA dataset and Figure 1b, represents the details of resultant MFV.

Algorithm 1 Creates a Master Feature Vector with an intention to impute and remove missing values with a certain threshold to improve classification and feature selection and extraction process performance on the obtained LGA dataset L_0

Input: LGA dataset L_0 with f_0 features and n_0 records.

Output: LGA dataset L with f features and n records

- 1: For each r_{th} row in L_0 , add c_{th} classification column c , following [27] infant classification guidelines.
 - 2: Discretize each f_0 -th feature of L_0 with literature and pediatrician's expertise.
 - 3: Impute nan in v_{th} missing value of each r_{th} record of L_0 .
 - 4: Remove r_{th} record from L_0 with missing threshold of 10% from controls and 15% from the cases
 - 5: Impute discrete v_{th} value with mode of every f_0 -th feature.
 - 6: return LGA dataset L with f features and n records.
-

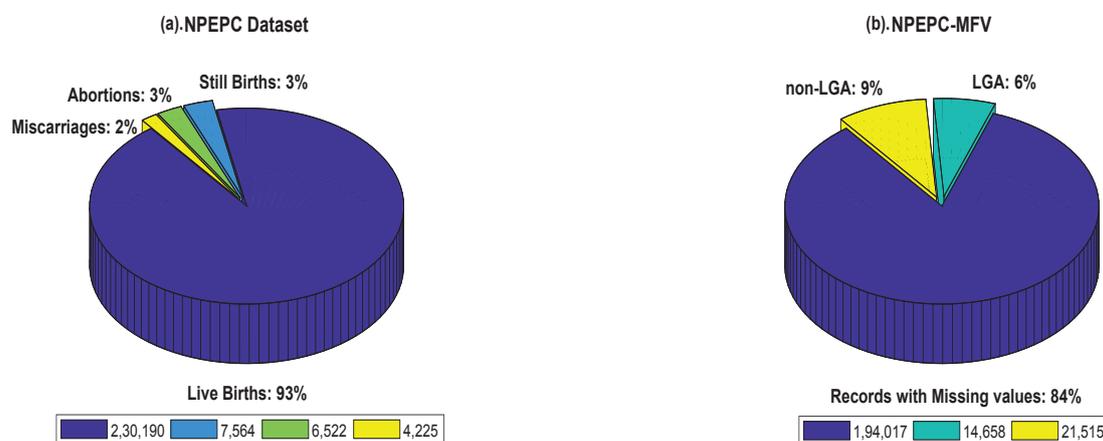


Figure 1. The details of the National Pregnancy and Examination Program of China dataset before and after applying Master Feature vector (MFV) creation algorithm where (a) represents the details of original Large for gestational age infants dataset and (b) represents the processed dataset following MFV creation algorithm.

3.3. Preparation of the Principal Feature Vector

An accurate and robust classification system requires discriminative features with reduced dimensions. Irrelevant and unnecessary features not only affect classifier performance but also demand excessive computational resources and time for the classification task [28–31]. A variety of feature selection, extraction, and reduction schemes are proposed by various researchers to deal with the curse of irrelevant and dimensionality problem of a classification system [23,28,32–34]. In this article,

we recommend using an ensemble of feature selection and extraction schemes to build an accurate and state of the art LGA prediction model. The development of Principal Feature Vector (PFV) is comprised of two different aspects. In the first aspect, GridSearch-based RFECV + IG feature selection scheme is applied to select, rank, and remove noisy features from the LGA dataset; whereas the second aspect further extracted features (i.e., for the sake of dimensionality reduction to eliminate generalization errors) obtained with the GridSearch-based RFECV + IG scheme using stack generalization to further improve the classification performance of the proposed scheme with lesser computational overhead. In the subsequent subsections, these schemes are precisely discussed.

3.3.1. GridSearch-Based RFECV + IG Feature Selection Scheme

Recursive Feature Elimination (RFE) is a scheme that excludes features based on its irrelevancy and low data integrity to a specified class distribution [25,35]. The elimination process is a continuous process until a complete list of deterministic feature subset is reached. The elimination process takes a classification model, and based on the classification weights, selects the weighted features with the elimination of noisy features. In addition, for a better elimination and selection process, the parameters for the classification model are required to be tuned. In this research, GridSearch, a popular technique for parameter tuning, is also accompanied by RFE to improve classification performance. Moreover, in the case of Recursive feature elimination with cross-validation (RFECV), the fitting is accompanied by testing, it uses training and test splits provided with a given folding parameter that helps in minimizing generalization errors.

Support Vector Machine (SVM) with linear and rbf kernels [36,37], Logistic Regression (LR) [38], and Decision Tree (DT) [39] classifiers are used in RFECV feature selection scheme with the parameters tuned with GridSearch with five-fold cross validation. During the tuning of GridSearch, for SVM (with linear kernel) and LR classifiers, 'C' is tuned in the range of 2^{-8} to 2^8 ; whereas, SVM with rbf kernel, 'C' is tuned in the range of 2^{-8} to 2^8 and gamma is tuned in the range of [1, 0.1, 0.01, 0.001, 0.00001, 10]. DT attributes are tuned for maximum depth: [1, 2, 3, 4, 5], criterion: ['gini', 'entropy'], and maximum features: ['sqrt', 'log2', 'auto']. The complete result of the selected features with discussed LGA classifier are presented in Table 2. All of the corresponding feature subsets are given to specified machine learning classifier for the establishment of first experimental setup.

Table 2. Features selected with GridSearch-based Recursive Feature Elimination with Cross-validation feature selection scheme with execution time using various machine learning classifiers for tuning parameters.

RFECV + Machine Learning Classifier	Selected Features	Time (s)
SVM (Linear kernel)	53	25537
SVM (RBF kernel)	99	201331
Logistic Regression	38	40386
Decision Tree	270	118

Furthermore, the IG feature selection scheme is used as an ensemble of feature selection process to rank above-induced feature subsets which are discussed as follows,

Information Gain (IG): is an extensively used feature selection scheme in a variety of machine learning problems, especially related to the medical domain. IG feature selection scheme works with an objective of uncertainty reduction in a feature vector. Once the level of uncertainty is known, the larger the uncertainty can be reduced, and more the information the feature can bring to the classification system. Thus, ultimately, we have a larger information gain that is brought to the system for the development of an efficient LGA classification model. In IG feature selection scheme, "information entropy" is used to measure the amount of information which is later calculated as the difference between dataset (B)'s information entropy including with and without LGA features x_i .

Furthermore, while training dataset B with I class labels, $\varrho(B)$ represents the information entropy of the LGA class distribution in B, which can be expressed as follows,

$$\varrho(B) = - \sum_{i=1}^I p_i \log_2 p_i \tag{1}$$

where P_i represents the probability of i-th class in the training dataset B. Moreover, x_i features with D distinct values can be used to partition dataset B into D distinct groups. Then, each group $B_d (d = 1, \dots, D)$ entropy is calculated as,

$$\varrho(B_d) = - \sum_{i=1}^I p_{di} \log_2 p_{di} \tag{2}$$

where p_{di} is the probability of i-th class in the training data subset B_d . Based on the fact that each subset may contain a different number of samples, i.e., each subset B_d contains Z_d samples where ($d = 1, \dots, D$), its weight is set to Z_d/Z . Information Gain with features x_i to partition dataset B can be written as

$$Information_{Gain}(B, x_i) = \varrho(B) - \sum_{d=1}^D \frac{Z_d}{Z} \varrho(B_d) \tag{3}$$

Based on calculated information gain of every attribute, the attributes with the highest IG are ranked with descending order for the further experimentations.

3.3.2. Feature Extraction and Dimension Reduction with Stacked Generalization

Stacked generalization (SG): also known as Stacking is a process that combines multiple classifiers to form an efficient classification system. It was introduced by Wolpert in 1992 [40]. The stacking process involves output generated from level-0 (base-level) classifiers as an input to level-1 (meta-level) classifier to improve classification performance using the process of cross validation is as follows,

Let us suppose that L is the obtained LGA dataset with a_i attributes with an associated y_i -th class label. Thus, $L = \{(a_i, y_i), i = 1, \dots, n\}$ refers to the level-0 of the LGA dataset. Based on K-fold cross-validation L is divided into K disjoint parts of L_1, L_2, \dots, L_k where at each k -th fold L_k is used as test and $L^{(-k)} = L - L_k$ is used as the test part. Later, N learning algorithms A_1, A_2, \dots, A_N are applied at training part $L^{(-k)}$ to build N level-0 classifiers C_1, C_2, \dots, C_N . The resultant concatenated predictions of each k -th fold at L_k of N level-0 classifiers with the actual class label are used to form meta-level vector (ML_k). It will be use during the establishment of level-1 classification.

With the development of complete meta-level vector (ML_k) also called level-1 data which is obtained by the union of each of the ML_k , where $k = 1, 2, 3, \dots, K$ during the cross-validation process, we applied the algorithm A_m to form the meta-level classifier C_m . During the development of C_m , the A_m could be any of the A_1, A_2, \dots, A_N or a different one. Based on this procedure, it is foreseen that after forming the meta-level data, the entire data is trained using the learning algorithms A_1, A_2, \dots, A_N to build final base-level classifiers C_1, C_2, \dots, C_N .

Ting et al. [41] proposed to use class probabilities instead of just using class labels for the formation of the meta-level feature vector, as it can better improve classification performance with an improved learning rate. Therefore, to classify a new instance predicted probabilities and predicted class labels by all level-0 classifiers are concatenated to form a meta-level classifier which has N components. Based on this formed meta-level feature vector, level-1 classifiers assign an actual class label to the final classification result of the input instance x . Figure 2. refers to the process of creation of K-fold cross-validation (the left part of the figure) while the right part of the figure represents the stacking process.

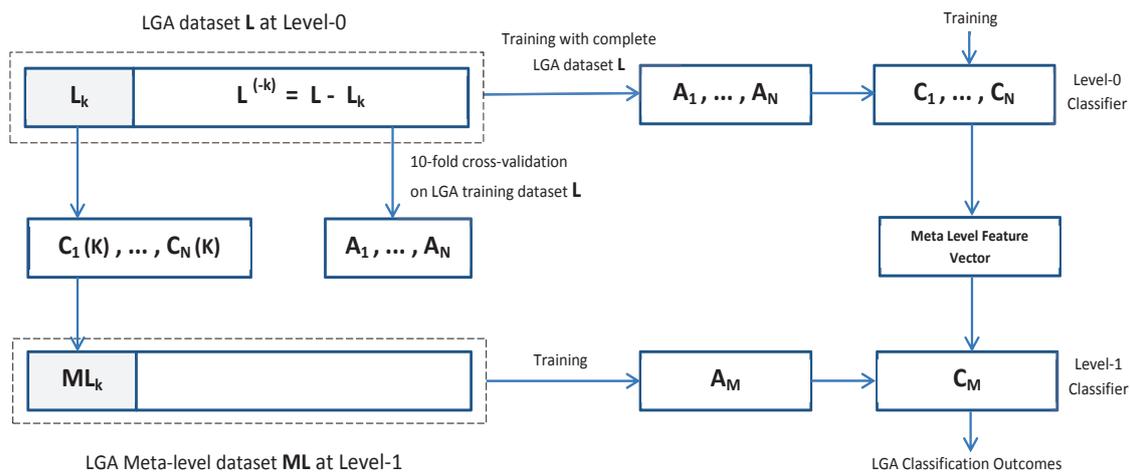


Figure 2. The establishment of K-fold cross validation process to create meta-level training data and stacking procedure with minimal generalization errors.

Feature Extraction and Dimension Reduction with Stacking: represents the above-defined process where the outputs of level-0 classification are combined to form complete meta-level features. It is, in fact, a process to extract discriminant features rather than just classification predictions. Previously, the combination strategy has already benefited different researcher to improve generalization errors during the classification task. On the basis of this, considering the best classification schemes on the said GridSearch-based RFECV + IG feature subsets, we proposed to use Logistic regression (LR) and Random forest (RF) classifiers for the creation of the discriminant meta-level feature subset followed by Support Vector Machine (SVM) classifiers at level-1 to establish a state of the art LGA classification system. The reason to chose LR, RF, and SVM is because of their efficient performance during the sensitivity analysis process.

Furthermore, before starting the feature extraction process, we proposed to use the below-defined technique to reduces the size of data to expedite classification speed and performance without further deletion of valuable records. Following [26], where the authors trained classifiers on a hyperspectral data (shape features data and magnitude features data) and combined their results with stacking to form a new feature subset that is extracted with level-0 classifiers prediction probabilities, actual and predicted outputs. Based on this, we subdivided the whole LGA dataset L , which is obtained as a result of MFV creation wherein total 36,172 (LGA = 14,658, non-LGA = 21,514) records are selected. We subdivided (LGA and non-LGA) / 2 and formulated a new LGA dataset which contains two equal number of records, and we call it subsets 1 and 2 of LGA dataset L . Each subset of L contains 18,086 (7329 = LGA, 10,757 = non-LGA)records. These two subsets are used at the level-0 of the stacking process, which is intensively discussed in the previous subsection. The complete process of feature extraction and classification task is presented in Figure 3 where at level-0 of stacking process RF and LR classifiers with ten-fold cross-validations are used for feature extraction task and SVM with the linear kernel is used at level-1 of the stacking process for the classification prediction task. These classifiers are selected because of their efficient performance on the said dataset L in the first group of experiments.

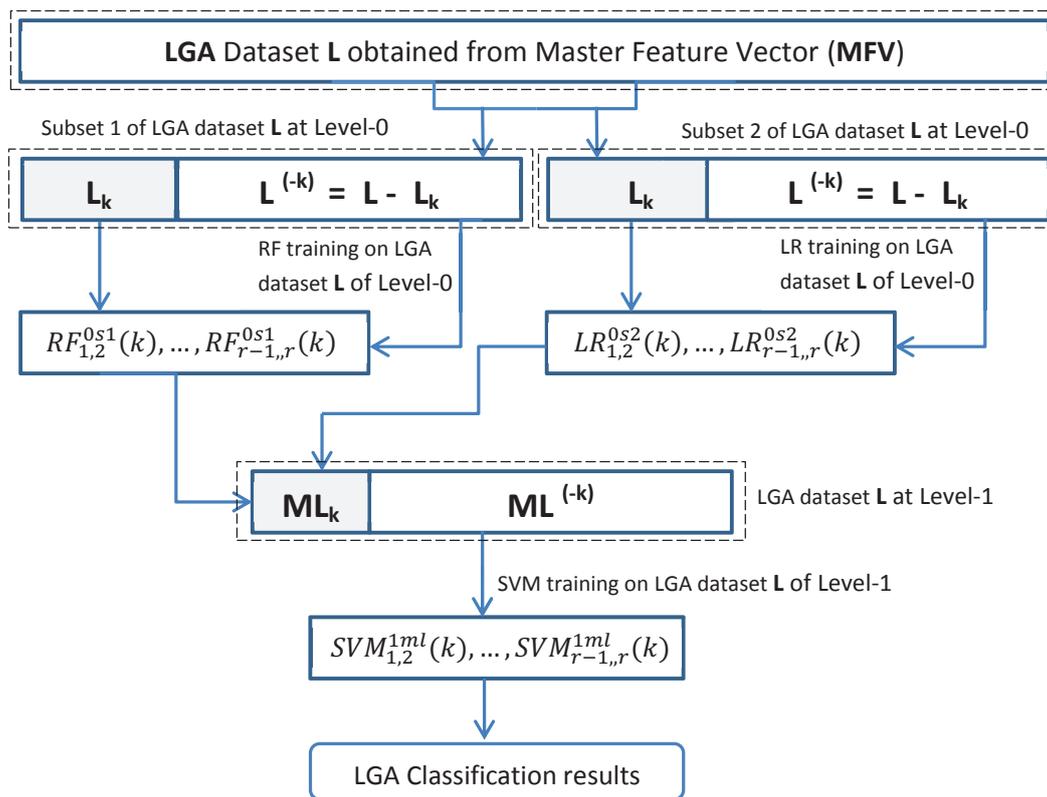


Figure 3. The complete classification procedure of proposed stacking scheme with the creation of meta-level training dataset at level-0 and level-1 of the stacking process.

3.4. LGA Classification Tools and Schemes

IBM SPSS statistics 22.0 and python are used for primary data processing. The LGA classification schemes are coded in python using the sci-kit-learn toolkit. Based on recent studies, four machine learning classifiers are selected. Logistic Regression (LR), Support Vector Machine (SVM) [42,43], and Random Forest (RF) classifiers are selected because of its outstanding performance in previously reported literature [22–24] and Decision Tree (DT) classifier is selected because of its simplicity, implicit feature screening process, easiness of data interpretability, and it also does not require any assumptions of linearity in the data [39]. In addition, SVM with RBF kernel is also exploited to observe the efficiency of SVM using its kernel trick; other kernels are not exploited because of its high computational time and cost.

3.5. Performance Evaluation Metrics

To evaluate the performance of proposed GridSearch-based RFECV feature selection scheme which is followed by IG feature selection scheme with stacked generalization, we selected precision, recall, accuracy, AUC, specificity, and F1 scores as a performance evaluation measures [44]. The possible outcomes of the proposed Gridsearch based RFECV + IG and Gridsearch based RFECV + IG followed by Staking can be described as

- TP (True Positive): Records correctly diagnosed as LGA or non-LGA;
- FP (False Positive): Records incorrectly diagnosed as LGA or non-LGA;
- TN (True Negative): Records correctly rejected by the classifier;
- FN (False Negative): Records incorrectly discarded by the classifier;

Furthermore, the derivation of these metrics are as follows,

Precision is the prediction, when it predicts yes and how often it is accurate or the number of true positive divided by the summation of a true positive and false positive.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Recall is the fraction of true positive and total actual positives in the dataset or the ability of the system to extract all relevant cases from the dataset or the number of true positive divided by the summation of a true positive and false negative.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

Accuracy is the correctness of the LGA classifiers in predicting LGA or non-LGA or the fraction of summation of true positive and true negative with the summation of true positive, true negative, false positive and false negative.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

AUC is used to analyse the correctness of the classification system in predicting a specific class. In fact, represents the classwise occupied area of a specific class.

$$AUC = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (7)$$

Specificity is the proportion of actual negatives which are correctly identified as it is, It represents true negative rate or the number of true negatives divided by the summation of a true negative and false positive.

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

F1 Score is the weighted average recall (true positive rate) and precision or it is the harmonic mean of recall and precision. Its formation is as follows.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (9)$$

4. Experiment Results

The experimental process is consolidated into two main processes. Where in the first process, two groups of experiments are performed. In the first group of experiments, RFECV feature selection scheme whose parameters are tuned with GridSearch is given to four widely used ML classifiers (Decision tree (DT), Logistic regression (LR), Support vector machine (SVM) with linear and RBF kernels) to classify LGA infants. In the second group of experiments, we used Information Gain (IG) feature selection scheme on previously identified features subsets identified using RFECV feature selection schemes. It is a sort of ensemble of the feature selection process where two feature selections schemes are added subsequently to remove noisy features with the identification of ranked feature subsets. Moreover, in the second process, stacking is proposed to extract features from the previous ensemble feature subsets to add another ensemble layer on the feature subsets to remove classifiers generalization errors and to improve classification performance. The results are of the experiments are presented in the following subsections.

4.1. Results of GridSearch Based RFECV + IG Feature Selection Scheme for LGA Prediction

To highlight the importance of proposed GridSearch-based RFECV feature selection scheme which is followed by IG feature selection scheme, we executed the initial experiments considering all features

as per MFV created and the features selected by GridSearch + RFECV features selection scheme. Table 3 can be referred for the details of the results. From the results, it is foreseen that the GridSearch + RFECV features selection scheme improved the LGA classification prediction scores and best performed with SVM classifier (using linear and RBF kernel). Furthermore, the classification performance of all of the classifiers on GridSearch + RFECV features subsets are also improved compared to the results of MFV features subset. Based on foreseen improvement, considering the primary objective of this research, which is to identify principal features subset for a better LGA prognosis, we executed the first proposed experimental process. Figure 4 presents the results of the initial experiment process where an ensemble of feature selection scheme is created using GridSearch + RFECV and IG feature selection scheme. From the results, it is discerned that all of the proposed classifiers outperformed with principal ten feature subsets. SVM (with the linear kernel) outperformed among all with prediction precision, recall, accuracy, AUC, specificity, and F1 scores of 0.97, 0.61, 0.83, 0.87, 0.999, and 0.74 respectively, followed by SVM (with RBF kernel), LR, and DT. SVM (with RBF kernel) and LR classifiers remained almost similar by producing similar performance metrics scores whereas DT classifier remained weak in producing noticeable performance metrics scores.

Table 3. Results of all features subsets selected by GridSearch-based RFECV features selection scheme using well known ML classifiers with 10-fold cross validation.

Scheme	Feature Subset	Metrics	SVM (Linear)	SVM (rbf)	Logistic Regression	Decision Tree
Master Feature Vector	All	Precision	0.8352	0.2025	0.8289	0.4970
	All	AUC	0.8447	0.2014	0.8281	0.5907
	All	Recall	0.6560	0.1198	0.6569	0.6991
	All	F1-Score	0.7166	0.1117	0.7236	0.5746
GridSearch + RFECV	GridSearch + RFECV(All)	Precision	0.9498	0.9691	0.9200	0.4961
	GridSearch + RFECV(All)	AUC	0.8690	0.8606	0.8659	0.5899
	GridSearch + RFECV(All)	Recall	0.6461	0.6059	0.6686	0.7008
	GridSearch + RFECV(All)	F1-Score	0.7663	0.7433	0.7716	0.5745

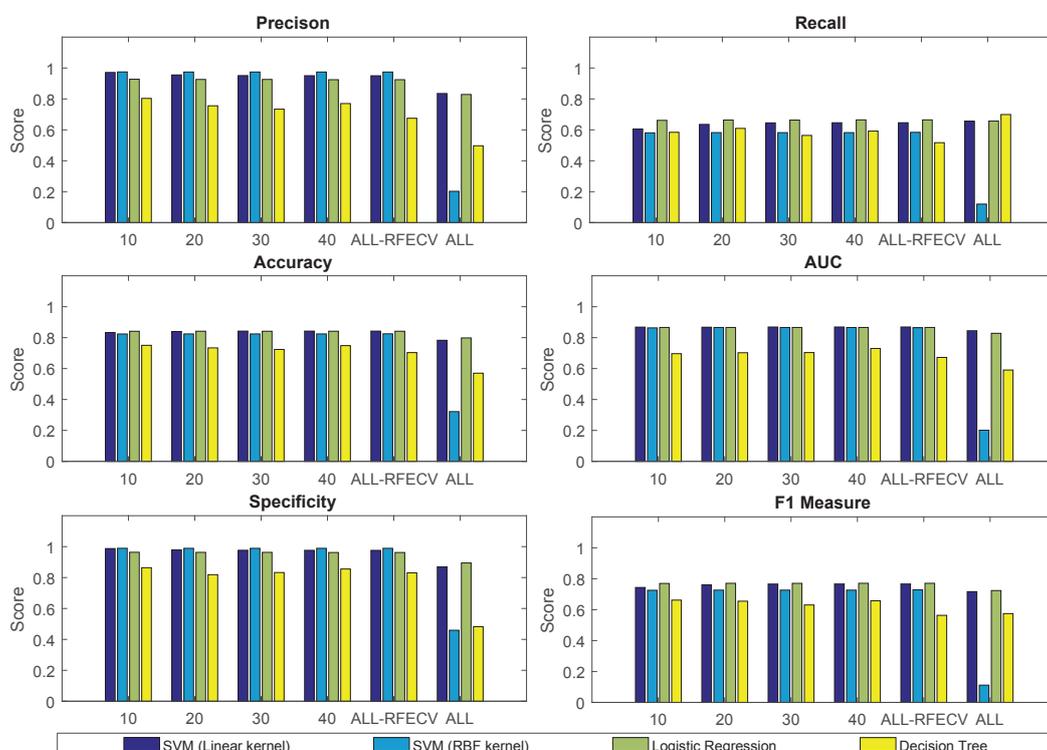


Figure 4. Comparative results of various machine learning classifiers on different features subset obtained as the result of applying GridSearch (for parameters tuning), RFECV feature selection scheme, and Information Gain (IG) feature selection scheme.

4.2. Results of GridSearch Based RFECV + IG Feature Selection Scheme with Stacking for LGA Prediction

To improve classification performance by removing generalization errors of the selected classifiers, we executed the second experimental process. The objective is to reduce or eliminate generalization errors with expedited classification performance using stacking, where level-0 of stacking is used for principal feature extraction with the intention of dimension reduction and level-1 is used to remove generalization errors to improve classification performance. Figure 5 presents the complete results of the proposed scheme. From the results, it is evident that the performance metrics scores are improved drastically, but with principal ten feature subsets, the results are noticeable. SVM (linear kernel) remained best with prediction precision, recall, accuracy, AUC, specificity, and F1 scores of 0.92, 0.87, 0.92, 0.95, 0.95, and 0.89 respectively, followed by SVM (RBF kernel), LR, and DT. SVM (RBF kernel) and LR classifiers remained almost similar by producing similar performance metrics scores whereas DT classifier remained weak in producing noticeable performance metrics scores.

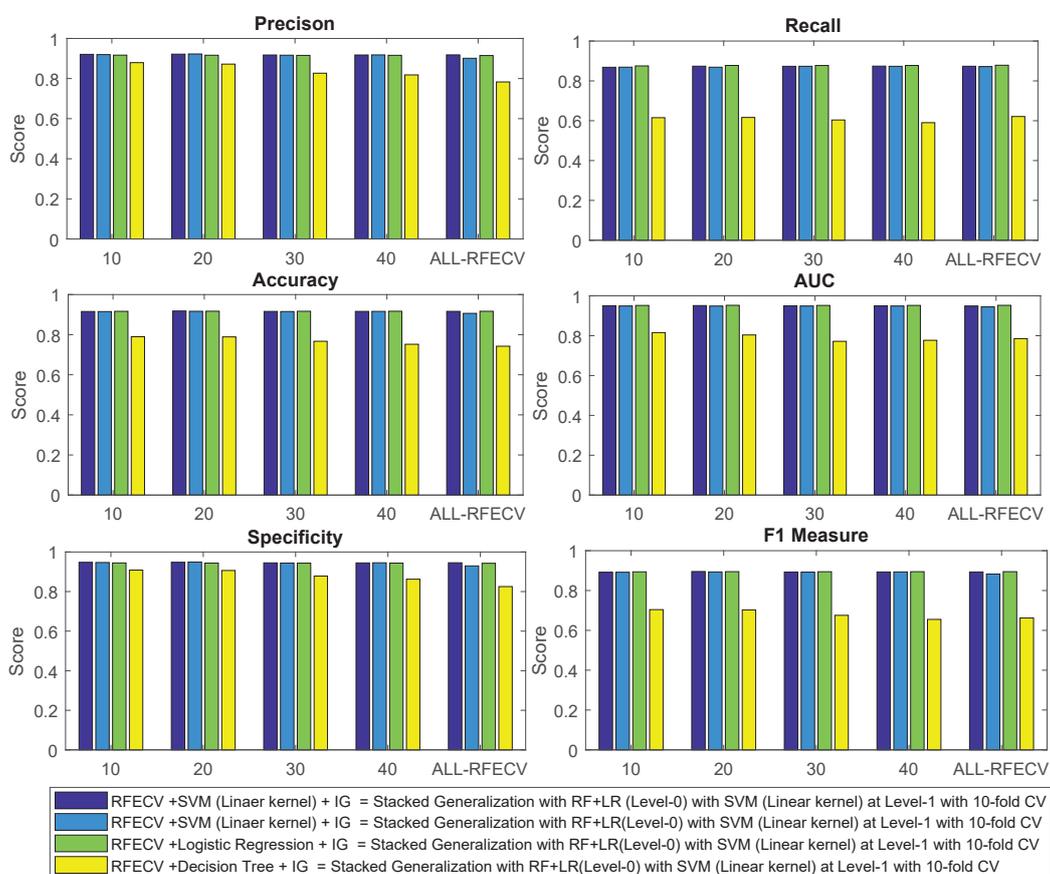


Figure 5. Comparative results of various stacked-classifiers on different features subset obtained as the result of applying GridSearch (for parameters tuning), RFECV feature selection scheme, Information Gain (IG) feature selection scheme for ranking of features, and Stacking to extract new features and to eliminate classifiers generalization errors.

5. Discussions and Comparative Analysis with Existing State-of-the-Art LGA Classifications

The proposed scheme for the classification of LGA fetus using stacked generalization with an ensemble feature selection scheme proved best in the selection of useful features subset which can accurately identify a fetus with its gestational parameters. From the results, it is also evident that the ranked ten principal features subset by every feature selection scheme remained best among all feature subsets and produced its highest prediction performance metrics scores. Table 4 presents the comparative best results of all three group of experiments with the proposed ensemble of feature

selection and extraction techniques with stack-generalization. From the results, it is observed that among the three experimentations, SVM (linear kernel) classifiers outperformed with the production of highest prediction performance metrics scores and remained best with prediction precision, recall, accuracy, AUC, specificity, and F1 scores of 0.92, 0.87, 0.92, 0.95, 0.95, and 0.89 respectively. The reason for being the best is because of the formation of maximum hyperplane between the LGA and non-LGA class. The creation of maximum hyperplane is possible because of easily superable feature subsets induced as the result of applying the proposed ensemble of feature selection and extraction schemes with stacked generalization. SVM (RBF kernel) classifier is also suitable for LGA classification task because of its impressive results but is not recommended because of its computational complexity. Furthermore, LR classifier can also be used for the said classification task, but DT classifier is never recommended due to its low performance. The reason for DT classifier for being insignificant might be because of inadequacy in applying regression and possibility of duplication with the same sub-tree on different paths while predicting values.

Table 4. Comparative best results of all three group of experiments with the proposed ensemble of feature selection and extraction techniques with stack-generalization.

Experiment Type	Best Classifier	Size	Precision	AUC	Recall	Accuracy	Specificity	F1
GirdSearch with tuned parameters	SVM (linear kernel)	All	0.949	0.843	0.646	0.842	0.976	0.766
GirdSearch + RFECV + Information Gain	SVM (linear kernel)	10	0.971	0.868	0.606	0.833	0.987	0.744
GridSearch + RFECV + IG + Stack generalization	Stacked SVM (linear kernel)	10	0.920	0.950	0.8683	0.9156	0.9478	0.8921

The significance of the proposed scheme is highlighted by comparing the results of the proposed scheme with existing state-of-arts LGA classification schemes. Table 5 presents the comparative best results of recently published schemes on the same dataset with the proposed scheme. The results reveals that the highest prediction performance metrics scores (i.e., precision = 0.92, AUC = 0.95, recall = 0.87, accuracy = 0.92, specificity = 0.95, and, F1 = 0.89) are obtained by the proposed scheme with SVM (linear kernel) using ten principal features subset. Table 6 present the results of ranked ten principal feature subset of GridSearch-based RFECV + IG feature selection scheme with four ML classifiers using ten-fold cross-validation. From the comparative analysis of the results of the proposed scheme, it is also discerned that the feature engineering and classification schemes of this research best suits the process of establishing a state-of-art LGA prognosis process with improved classification performance with lesser computational overhead. The reason for the improvement in classification performance is because of the extraction of reduced numbers of discriminant features subset, which eventually helps in removing LGA classifiers complexity with decreased generalization errors to improve LGA classification accuracy.

$$CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6 * N}} \quad (10)$$

Table 5. Comparative best results of proposed and previously published schemes on the said Large for gestational age dataset.

Baseline	Scheme	Precision	AUC	Recall	Accuracy	Specificity	F1
Akhtar et al. [22]	IG + ML Classifier	0.71	0.71	-	-	-	-
Akhtar et al. [23]	Proposed Ensemble Technique + ML Classifiers	0.85	0.72	-	-	-	-
Akhtar et al. [24]	Proposed Expert Driven + ML Classifiers	0.95	0.86	-	0.85	-	-
This Research	GridSearch + RFECV + IG + Stack Generalization	0.92	0.95	0.87	0.92	0.95	0.89

Table 6. The ranked ten principal feature subsets of GridSearch-based RFECV + IG feature selection scheme with four ML classifiers using ten-fold cross validation.

Number	GridSearch + RFECV + IG + SVM(Linear)	GridSearch + RFECV + IG + SVM(RBF)	GridSearch + RFECV + IG + LR	GridSearch + RFECV + IG + DT
1	Pregnancy History	Pregnancy History	Pregnancy History	Pregnancy History
2	Smoking (m)	Smoking (m)	Contraception Used	Contraception Used
3	Contraception Used	Toxic Pesticide	# Full Term Birth	Normal Birth
4	# Full Term Birth	Contraception Used	# of Pregnancies	# Full Term Birth
5	# of Pregnancies	# Full Term Birth	Evaluation Result	# of Pregnancies
6	Evaluation Result	# of Pregnancies	High Risk Fetus ?	Region Name
7	High Risk Fetus ?	Evaluation Result	Delivery Week	Follow-up Institution
8	Delivery Week	High Risk Fetus ?	Normal Birth	Delivery Week
9	Normal Birth	Delivery Week	Induced Labour	Child Birth Province
10	# of Fetuses	Premature Delivery	# of Fetuses	Child Birth Town

Moreover, Friedman and Bonferroni–Dunn tests are also introduced to rank and highlight the significance difference between the results reported in Figures 4 and 5. Initially, Friedman test considering ($p < 0.05$) is employed to rank the classifiers based on the result of said experiments. The longitudinal axis in Figures 6 and 7 represents the average mean ranking calculated by using Friedman test on all group of experiments. From the results it is foreseen that SVM with linear kernel outperformed in almost each group of experiment. In addition, Bonferroni–Dunn test is also employed in the significance level of $\alpha < 0.05$, $\alpha < 0.01$, and $\alpha < 0.001$ to the results of Friedman test. Equation (10) is used to calculate Critical Distance (CD) used in Bonferroni–Dunn Test. Based on provided guidelines by the author of [45], for Figure 6, we selected $N = 6$, and $k = 4$, q = is equal to $q_\alpha(0.05) = 3.4077$, $q_\alpha(0.01) = 4.089$, and $q_\alpha(0.001) = 4.9198$ whereas, for Figure 7, $N = 5$, and $k = 4$, q = is equal to $q_\alpha(0.05) = 3.3045$, $q_\alpha(0.01) = 4.004$, and $q_\alpha(0.001) = 4.8444$. Based on these figures’ results, it is observed that SVM has the largest difference in-between pair-wise means of control group with the critical values which validates the previously concluded remarks of using ranked ten features subset with SVM (linear kernel) classifier as an important mean to diagnose infants as LGA or non-LGA.

Furthermore, the proposed scheme has the potential to classify various disease classes accurately using gestational parameters as suggested by the panel of experts of different domains. The limitation of the proposed scheme is that it experiments only for LGA dataset. However, it has the potential to produce accurate results for Small for Gestational Age (SGA) infants as well, which we will explore in our future work. In addition, as previously discussed that machine learning techniques on LGA have never been exercised extensively, so this research presents an extensive work that can facilitate paediatricians and researcher to extend their research in the defined area. Moreover, in our future work, deep learning techniques, such Standard Deep Neural network (NN) [46], Hierarchical deep learning (HDL) [47], Random multimodel deep learning (RMDL) or deep perceptron [48] will also be exploited to add more scientific results to the related domain.

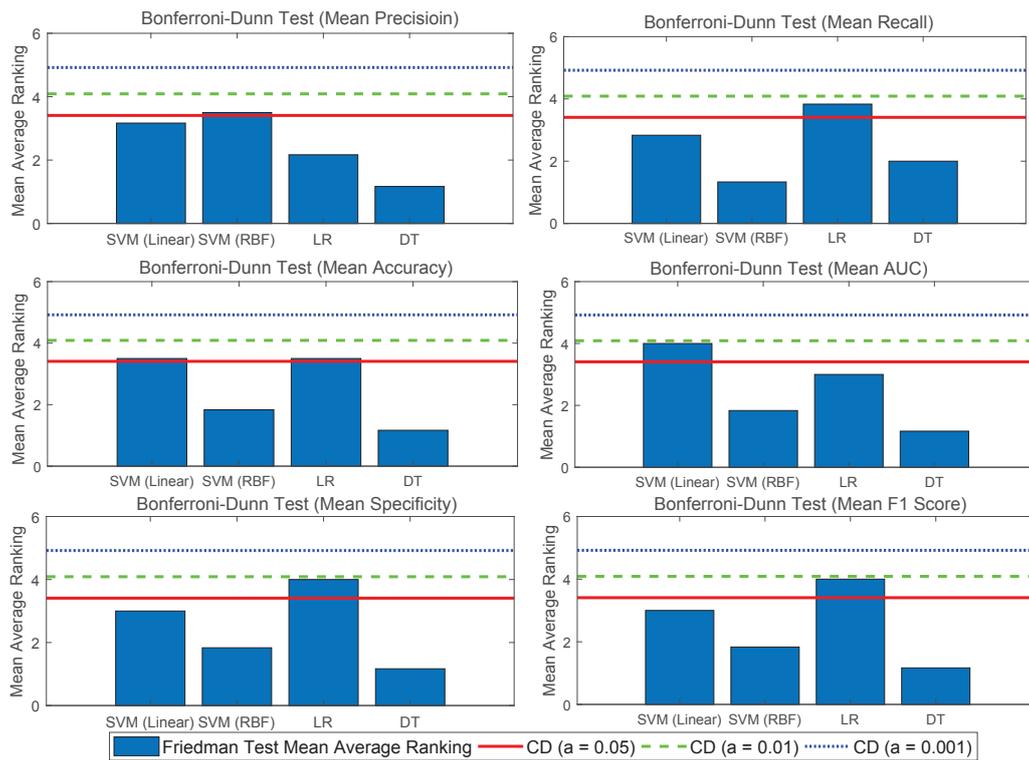


Figure 6. Result ranked with Friedman test and Bonferroni–Dunn test of four ML algorithms (SVM (Linear kernel), SVM (RBF kernel), LR, and DT) with precision, recall, accuracy, AUC, Specificity, and F1 Score in significant levels of $\alpha < 0.05$, $\alpha < 0.01$, and $\alpha < 0.001$ taking DT as a control algorithm in Figure 4 results.

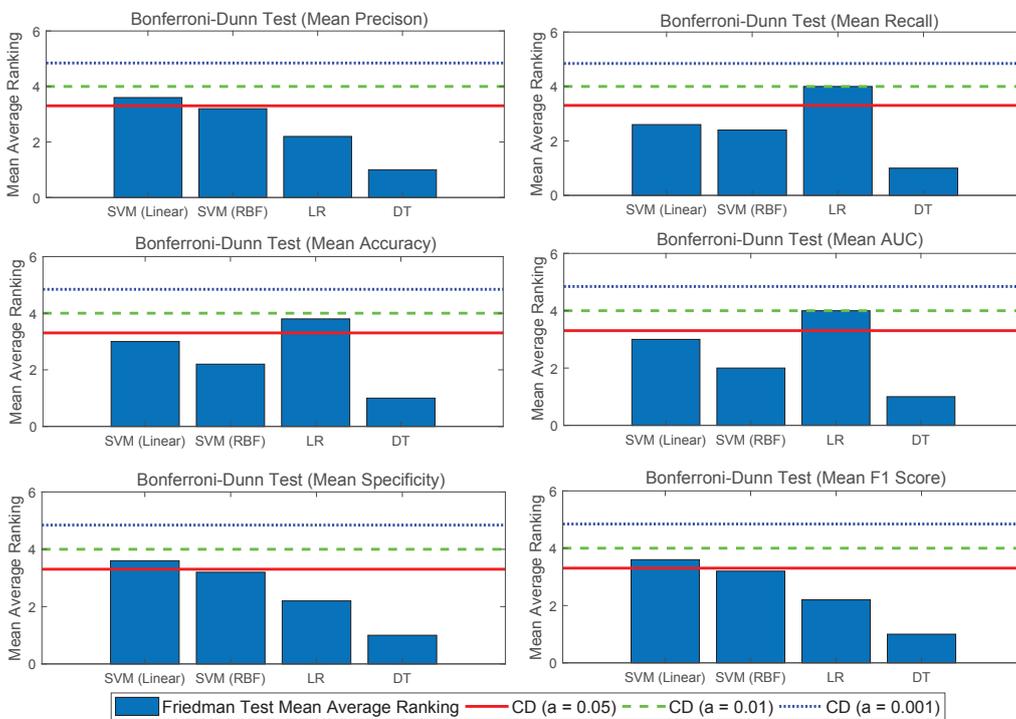


Figure 7. Result ranked with Friedman test and Bonferroni–Dunn test of four ML algorithms (SVM (Linear kernel), SVM (RBF kernel), LR, and DT) with precision, recall, accuracy, AUC, Specificity, and F1 Score in significant levels of $\alpha < 0.05$, $\alpha < 0.01$, and $\alpha < 0.001$ taking DT as a control algorithm in Figure 5 results.

6. Conclusions and Future Work

In this research, an LGA classification model is developed to classify a fetus as LGA or non-LGA. It is composed of the GridSearch-based RFECV + IG feature selection scheme followed by stacking to select, rank, and extract significant features from the LGA dataset. The proposed LGA classification scheme using stacking with an ensemble of feature selection and extraction schemes yielded better performance in terms of precision, AUC, recall, accuracy, specificity, and, F1 scores, when it is compared with existing state-of-the-art schemes. This study helps to establish a comprehensive comparison of various decision models performance on the said LGA dataset, which concludes that GridSearch based RFECV+IG feature selection scheme with stacking using SVM (linear kernel) best suits the said classification process followed by SVM (RBF kernel) and LR classifiers. DT classifier is not suggested because of its low performance. Almost every classification scheme best performed with ten principal feature subsets. It is evident from the results that the proposed scheme has the potential to classify an LGA fetus accurately and efficiently. In addition, the promising results indicate that the paediatrician and experts can use the proposed model for the establishment of an efficient LGA classification system as a second opinion, which has the potential to assist them in establishing a proper LGA prognosis process with ranked features subset. In the future, the proposed scheme will also be extended for the classification and identification of Small for Gestational Age (SGA) infants with better performance metrics scores and deep learning techniques will also be exploited to improve classification performance.

Author Contributions: F.A., J.L., and Y.P. conceived and designed the experiments; F.A. Performed the experiments; F.A., A.I., A.R., M.A. analysed the data; F.A., J.L., Y.P., and Q.W. contributed reagents/materials/analysis tools; F.A. wrote the paper.

Funding: This study is supported by the Beijing Nature Science Foundation of China (Z160003).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chiavaroli, V.; Castorani, V.; Guidone, P.; Derraik, J.B.; Liberati, M.; Chiarelli, F.; Mohn, A. Incidence of infants born small- and large-for-gestational-age in an Italian cohort over a 20-year period and associated risk factors. *Ital. J. Pediatr.* **2016**, *42*, 42. [[CrossRef](#)]
2. Mendez-Figueroa, H.; Truong, V.T.; Pedroza, C.; Chauhan, S.P. Large for gestational age infants and adverse outcomes among uncomplicated pregnancies at term. *Am. J. Perinatol.* **2017**, *34*, 655–662. [[PubMed](#)]
3. Battaglia, F.C.; Lubchenco, L.O. A practical classification of newborn infants by weight and gestational age. *J. Pediatr.* **1967**, *71*, 159–163. [[CrossRef](#)]
4. Lazer, S.; Biale, Y.; Mazor, M.; Lewenthal, H.; Insler, V. Complications associated with the macrosomic fetus. *J. Reprod. Med.* **1986**, *31*, 501–505. [[PubMed](#)]
5. Meshari, A.A.; Silva, S.D.; Rahman, I. Fetal macrosomia, maternal risks and fetal outcome. *Int. J. Gynecol. Obstet.* **1990**, *32*, 215–222. [[CrossRef](#)]
6. Boney, C.M.; Verma, A.; Tucker, R.; Vohr, B.R. Metabolic syndrome in childhood: Association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* **2005**, *115*, 290–296. [[CrossRef](#)] [[PubMed](#)]
7. Dyer, J.S.; Rosenfeld, C.R.; Rice, J.; Rice, M.; Hardin, D.S. Insulin resistance in Hispanic large-for-gestational-age neonates at birth. *Early Hum. Dev.* **2007**, *83*, S138. [[CrossRef](#)]
8. Ingrid, W.M.D.; Axelsson, O.; Bergstrom, R. Maternal factors associated with high birth weight. *Acta Obstet. Gynecol. Scand.* **2011**, *70*, 55–61.
9. Dietz, W.H. Overweight in childhood and adolescence. *N. Engl. J. Med.* **2004**, *350*, 855–857. [[CrossRef](#)]
10. Van Assche, F.A.; Devlieger, R.; Harder, T.; Plagemann, A. Mitogenic effect of insulin and developmental programming. *Diabetologia* **2010**, *53*, 1243–1243. [[CrossRef](#)]
11. Xu, H.; Simonet, F.; Luo, Z.C. Optimal birth weight percentile cut-offs in defining small- or large-for-gestational-age. *Acta Paed.* **2010**, *99*, 550–555. [[CrossRef](#)] [[PubMed](#)]

12. Zhang, S.; Wang, A.; Shen, H. Design implementation and significance of Chinese free pre-pregnancy eugenics checks projec. *Natl. Med. J. China* **2015**, *95*, 162–165.
13. Shen, Y.; Zhao, W.; Lin, J.; Liu, F. Accuracy of sonographic fetal weight estimation prior to delivery in a Chinese han population. *J. Clin. Ultrasound* **2017**, *45*, 465–471. [[CrossRef](#)] [[PubMed](#)]
14. Blue, N.R.; Jmp, Y.; Holbrook, B.D.; Nirgudkar, P.A.; Mozurkewich, E.L. Abdominal circumference alone versus estimated fetal weight after 24 weeks to predict small or large for gestational age at birth: A meta-analysis. *Am. J. Perinatol.* **2017**, *34*, 1115–1124. [[PubMed](#)]
15. Harper, L.M.; Jauk, V.C.; Owen, J.; Biggio, J.R. The utility of ultrasound surveillance of fluid and growth in obese women. *Am. J. Obstet. Gynecol.* **2014**, *211*, 524.e1–524.e8. [[CrossRef](#)]
16. Chen, Q.; Wei, J.; Tong, M.; Yu, L.; Lee, A.C.; Gao, Y.F.; Zhao, M. Associations between body mass index and maternal weight gain on the delivery of LGA infants in Chinese women with gestational diabetes mellitus. *J. Diabetes Its Complicat.* **2015**, *29*, 1037–1041. [[CrossRef](#)]
17. Moore, G.S.; Kneitel, A.W.; Walker, C.K.; Gilbert, W.M.; Xing, G. Autism risk in small- and large-for-gestational-age infants. *Am. J. Obstet. Gynecol.* **2012**, *206*, 314.e1–314.e9. [[CrossRef](#)]
18. Luangkwan, S.; Vetchapanpasat, S.; Panditpanitcha, P.; Yimsabai, R.; Subhaluksuksakorn, P.; Loyd, R.A.; Uengarporn, N. Risk factors of small for gestational age and large for gestational age at Buriram hospital. *J. Med. Assoc. Thai* **2015**, *98*, S71–S78.
19. Khanolkar, A.R.; Hanley, G.E.; Koupil, I.; Janssen, P.A. 2009 IOM guidelines for gestational weight gain: How well do they predict outcomes across ethnic groups. *Ethn. Health* **2017**, 1–16. [[CrossRef](#)]
20. Kominiarek, M.A.; Grobman, W.; Adam, E.; Buss, C.; Culhane, J.; Entringer, S.; Simhan, H.; Wadhwa, P.D.; Kim, K.Y.; Keenan-Devlin, L.; et al. Stress during pregnancy and gestational weight gain. *J. Perinatol.* **2018**, *38*, 462–467. [[CrossRef](#)]
21. Shepherd, E.; Gomersall, J.C.; Tieu, J.; Han, S.; Crowther, C.A.; Middleton, P. Combined diet and exercise interventions for preventing gestational diabetes mellitus. *Cochrane Libr.* **2017**, *11*. [[CrossRef](#)] [[PubMed](#)]
22. Faheem Akhtar, J.L.; Guan., Y. Monitoring bio-chemical indicators using machine learning techniques for an effective large for gestational age prediction model with reduced computational overhead. In Proceedings of the 7th International Conference on Frontier Computing (FC 2018) - Theory, Technologies and Applications, Kuala Lumpur, Malaysia, 3–6 July 2018.
23. Akhtar, F.; Li, J.; Azeem, M.; Chen, S.; Pan, H.; Wang, Q.; Yang, J.J. Effective LGA prediction using ML techniques monitoring biochemical indicators. *J. Supercomput.* **2019**. [[CrossRef](#)]
24. Akhtar, F.; Li, J.; Pei, Y.; Azeem, M. A semi-supervised technique for IGA prognosis. In *Proceedings of The International Workshop on Future Technology FUTECH 2019*; Korean Institute of Information Technology: Daejeon, Korea, 2018; pp. 36–37.
25. Park, D.; Lee, M.; Park, S.; Seong, J.K.; Youn, I. Determination of optimal heart rate variability features based on SVM-recursive feature elimination for cumulative stress monitoring using ECG sensor. *Sensors* **2018**, *18*, 2387. [[CrossRef](#)] [[PubMed](#)]
26. Chen, J.; Wang, C.; Wang, R. Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2193–2205. [[CrossRef](#)]
27. Zhu, L.; Zhang, R.; Zhang, S.; Shi, W.; Yan, W.; Wang, X.; Lyu, Q.; Liu, L.; Zhou, Q.; Qiu, Q. Chinese neonatal birth weight curve for different gestational age. *Zhonghua Er Ke Za Zhi* **2015**, *53*, 97–103.
28. Li, J.; Liu, L.; Zhou, M.C.; Yang, J.J.; Chen, S.; Liu, H.T.; Wang, Q.; Pan, H.; Sun, Z.H.; Tan, F. Feature selection and prediction of small-for-gestational-age infants. *J. Ambient Intell. Humaniz. Comput.* **2018**, 1–15. [[CrossRef](#)]
29. Li, J.; Liu, L.; Sun, J.; Mo, H.; Yang, J.; Chen, S.; Liu, H.; Wang, Q.; Pan, H. Comparison of different machine learning approaches to predict small for gestational age infants. *IEEE Trans. Big Data* **2016**, 1–14. [[CrossRef](#)]
30. Yang, J.J.; Li, J.; Mulder, J.; Wang, Y.; Chen, S.; Wu, H.; Wang, Q.; Pan, H. Emerging information technologies for enhanced healthcare. *Comput. Ind.* **2015**, *69*, 3–11. [[CrossRef](#)]
31. Miao, J.; Niu, L. A survey on feature selection. *Procedia Comput. Sci.* **2016**, *91*, 919–926. [[CrossRef](#)]
32. Li, J.; Wang, F. Semi-supervised learning via mean field methods. *Neurocomputing* **2016**, *177*, 385–393. [[CrossRef](#)]
33. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [[CrossRef](#)]

34. Cunningham, J.P.; Ghahramani, Z. Linear dimensionality reduction: Survey, insights, and generalizations. *J. Mach. Learn. Res.* **2015**, *16*, 2859–2900.
35. Vapnik, V.N. *Statistical Learning Theory*; Springer: Berlin, Germany, 1998. doi:10.1007/978-1-4757-3264-1. [[CrossRef](#)]
36. Adankon, M.M.; Cheriet, M.; Biem, A. Semisupervised least squares support vector machine. *IEEE Trans. Neural Netw.* **2009**, *20*, 1858–1870. [[CrossRef](#)] [[PubMed](#)]
37. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
38. Bammann, K. Statistical models: Theory and practice. *Biometrics* **2006**, *62*, 943. [[CrossRef](#)]
39. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [[CrossRef](#)]
40. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
41. Ting, K.M.; Witten, I.H. Issues in stacked generalization. *J. Artif. Intell. Res.* **1999**, *10*, 271–289. [[CrossRef](#)]
42. Shmueli, A.; I. Nassie, D.; Hirsch, L.; Ashwal, E.; Wiznitzer, A.; Yogev, Y.; Aviram, A. 241: Prerecognition of large for gestational age (LGA) fetus and its consequences. *Am. J. Obstet. Gynecol.* **2017**, *216*, S150–S151. [[CrossRef](#)]
43. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT'92, Pittsburgh, PA, USA, 27–29 July 1992; ACM: New York, NY, USA, 1992; pp. 144–152. [[CrossRef](#)]
44. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
45. Zar, J.H. *Biostatistical Analysis*, 4th ed.; Pearson Education: Upper Saddle River, NJ, USA, 1999.
46. Eykholt, K.; Evtimov, I.; Fernandes, E.; Li, B.; Rahmati, A.; Xiao, C.; Prakash, A.; Kohno, T.; Song, D. Robust physical-world attacks on deep learning visual classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1625–1634.
47. Kowsari, K.; Brown, D.E.; Heidarysafa, M.; Jafari Meimandi, K.; Gerber, M.S.; Barnes, L.E. HDLTex: Hierarchical deep learning for text classification. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 364–371. [[CrossRef](#)]
48. Kowsari, K.; Heidarysafa, M.; Brown, D.E.; Meimandi, K.J.; Barnes, L.E. Rmdl: Random multimodel deep learning for classification. In Proceedings of the 2nd International Conference on Information System and Data Mining, Lakeland, FL, USA, 9–11 April 2018; pp. 19–28.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).