



A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies

Saurabh Tewari¹ · U. D. Dwivedi¹

Received: 19 August 2019 / Accepted: 16 January 2020 / Published online: 4 February 2020
© The Author(s) 2020

Abstract

Mudstone reservoirs demand accurate information about subsurface lithofacies for field development and production. Normally, quantitative lithofacies modeling is performed using well logs data to identify subsurface lithofacies. Well logs data, recorded from these unconventional mudstone formations, are complex in nature. Therefore, identification of lithofacies, using conventional interpretation techniques, is a challenging task. Several data-driven machine learning models have been proposed in the literature to recognize mudstone lithofacies. Recently, heterogeneous ensemble methods (HEMs) have emerged as robust, more reliable and accurate intelligent techniques for solving pattern recognition problems. In this paper, two HEMs, namely voting and stacking, ensembles have been applied for the quantitative modeling of mudstone lithofacies using Kansas oil-field data. The prediction performance of HEMs is also compared with four state-of-the-art classifiers, namely support vector machine, multilayer perceptron, gradient boosting, and random forest. Moreover, the contribution of each well logs on the prediction performance of classifiers has been analyzed using the Relief algorithm. Further, validation curve and grid search techniques have also been applied to obtain valid search ranges and optimum values for HEM parameters. The comparison of the test results confirms the superiority of stacking ensemble over all the above-mentioned paradigms applied in the paper for lithofacies modeling. This research work is specially designed to evaluate worst- to best-case scenarios in lithofacies modeling. Prediction accuracy of individual facies has also been determined, and maximum overall prediction accuracy is obtained using stacking ensemble.

Keywords Ensemble methods · Voting ensemble · Stacking ensemble · Reservoir characterization · Lithofacies modeling · Well log interpretations

Abbreviations

CL	Clay	HEM	Heterogeneous ensemble method
CM	Carbonate mudstone	HoEMs	Homogeneous ensemble methods
DM	Dolomitic mudstone	10-FCV	10-Fold cross-validation
DP	Dolomitic packstone	SG	Savitzky–Golay
DS	Dolomitic sandstone	SP	Spontaneous potential log
DW	Dolomitic wackstone	TT1	Acoustic transit time 2 log
PS	Packstone	DT	Acoustic transit time 1 log
SS	Siltstone	MN	Micronormal resistivity log
WS	Wackstone	MI	Microinverse resistivity log
MLP	Multilinear perceptron	MCAL	Caliper 2 log
SVM	Support vector machine	DCAL	Caliper 1 log
RF	Random forest	RLL3	Deep laterolog resistivity log
GB	Gradient boosting	RILM	Medium induction log
		RILD	Deep induction log
		RHOC	Density correction log
		NPOR	Neutron log
		GR	Gamma ray log
		DPOR	Density log

✉ U. D. Dwivedi
udwivedi@rgipt.ac.in

¹ Rajiv Gandhi Institute of Petroleum Technology, Amethi,
Uttar Pradesh 229304, India

SPOR Sonic log
viz. Namely

Introduction

Mudstones are widely occurring siliciclastic sedimentary rocks that behave as a source, cap and reservoir rock for hydrocarbon systems (Aplin and Macquaker 2011). It contains hidden organic-rich sweet spots and shale gas reservoirs which are favorable for petroleum production. Mudstone reservoirs are unconventional complex geological systems and provide challenges to conventional lithofacies interpretation techniques. Extraction of hydrocarbon from these unconventional resources requires accurate information of formation lithofacies, its association with petrophysical properties of reservoir rock and their spatial distribution (Spain et al. 2015). Conventionally, qualitative analysis is performed to recognize subsurface mudstone facies using core analysis, geomechanical spectroscopy logs, Rock–Eval pyrolysis, etc. (Bhattacharya et al. 2016). The conventional methodology is found to be inconvenient, tiresome, expensive in nature and requires high domain expertise.

Recognition of subsurface lithofacies is much researched topic and still a thought-provoking problem due to the uncertainty associated with reservoir measurements (Chaki et al. 2015; Bhattacharya et al. 2016). Quantitative modeling of lithofacies is essential to assess the potential of unconventional hydrocarbon reservoirs lying in mudstone formations. It also helps to understand the diagenetic and depositional burial history of these reservoirs (Aplin and Macquaker 2011). Quantitative information about underlying mudstone lithofacies can be taken out from conventional well logs which record the physical properties of rocks along with the reservoir depth. Several logging techniques such as wireline, logging-while-drilling, measurement-while-logging, etc., are utilized to generate wide varieties of well logs for measuring petrophysical properties of reservoir rock (Anifowose et al. 2015, 2017). Various researchers have reported that logging data are found to be nonlinear, high dimensional, complex and noisy in nature due to the heterogeneity and spatial distribution of reservoir properties (Kocerberber and Collins 1990; Chaki et al. 2015; Bhattacharya et al. 2016; Tewari and Dwivedi 2018a). Therefore, manual identification of geological lithofacies from sensory well logs is an impractical and tedious job even for expert field engineers. Thus, advance predictive machine learning models are suggested for extracting the lithofacies information from well logs.

Several machine learning models have been proposed to extract the facies information of conventional reservoir using well logs data. However, only few research works are available for unconventional mudstone reservoirs. Machine

learning paradigms utilized for quantitative lithofacies modeling of mudstone lithology are limited to unsupervised and supervised classifiers (Qi and Carr 2006; Ma 2011; Wang and Carr 2012; Anifowose et al. 2015; Bhattacharya et al. 2016). Aplin and Macquaker (2011) published a comprehensive review of mudstone lithology. They also studied the roles played by mudstone, viz. as an organic-rich source rock, cap rock and reservoir rock, from generation to storage of hydrocarbon. Li and Schieber (2017) did a detailed study about mudstone facies of the Henry Mountain Region of Utah. Qi and Carr (2006) employed an artificial neural network (ANN) model for the identification of carbonate lithofacies existing in Southwest Kansas from well logs data. Wang and Carr (2012) applied discriminant analysis, ANN, support vector machine (SVM) and fuzzy logic techniques for lithofacies modeling of the Appalachian basin at USA. They utilized core and seismic data along with well logs to develop a 3-D model of shale facies at the regional scale. Avanzini et al. (2016) implemented unsupervised cluster analysis for lithofacies classification to identify hidden productive sweet spots in the Barnett Shale formation. Bhattacharya et al. (2016) compared the performance of unsupervised and supervised machine learning models for mudstone facies present in Mahantango-Marcellus and Bakken Shale, USA. Bhattacharya et al. (2019) applied SVM for the identification of shale lithofacies of Bakken formation existing in North Dakota, USA. Table 1 contains a summary of important published research works related to mudstone lithofacies classification utilizing machine learning techniques.

All the above-mentioned research works are based on single supervised or unsupervised classifiers. However, it has been proved that the performance of single classifiers can be improved using hybrid computational models such as multiple classifier system, a committee of machines, composite systems, etc. (Dietterich 2000; Skurichina and Duin 2001). Multiple classifier system, like ensemble methods, can excavate more valuable information from raw sensory data. It combines the decisions of several classifiers together for classification and regression tasks. The ensemble approach can be categorized into two types: (a) homogeneous ensemble methods (HoEMs) such as bagging, random forest, rotational forest, random subspace, etc., and (b) heterogeneous ensemble methods (HEMs) such as voting, stacking, etc. HoEMs in feature space combine several hypotheses generated by the identical type of supervised classifiers which are utilized as base classifiers (e.g., a cluster of hundreds of SVMs). In the case of HEMs, different classifiers are utilized to generate and combine diverse hypotheses to achieve maximum possible prediction accuracy for existing feature space. It has been proved that heterogeneity in base classifiers helps to develop more reliable, robust and generalized classifier models (Sesmero et al. 2015). Ensemble methods have the capability to handle complex, nonlinear,

Table 1 Summary of some important published research works related to mudstone lithofacies classification utilizing machine learning techniques

S. no.	References	Methods	Number of wells, logs and facies	Field name
1.	Qi and Carr (2006)	ANNs	100 well data, 6 logs, 6 facies	Southwest Kansas
2.	Al-Anazi and Gates (2010)	Linear discriminant analysis (LDA), probabilistic neural network, general regression neural network and SVM	3 well data, 5 logs, 3 facies	Canada field
3.	Raeesi et al. (2012)	Competitive learning network and multilayer perceptron	4 well data, 2 logs, 4 facies	Farsi block, Persian Gulf
4.	Wang and Carr (2012)	LDA, fuzzy logic, ANNs and SVM	14 well data, 5 logs, 7 facies	Appalachian basin, USA
5.	Sebtosheikh and Salehi (2015)	SVM	5 well data, 9 logs, 3 facies	Iran
6.	Horrocks et al. (2015)	Naïve Bayes, ANNs and SVM	7 well data, 19 logs, 10 facies	Juandah East, Queensland, Australia
7.	Avanzini et al. (2016)	Ward's and K-means clustering algorithms	8 well data, 4 logs, 8 facies	Fort Worth Basin, Texas, USA
8.	Bhattacharya et al. (2016)	Self-organizing map, ANNs, SVM and multiresolution graph-based clustering	4 logs, 7 facies	North Dakota, USA
9.	Al-Mudhafar (2017)	Probabilistic neural networks	4 logs, 5 facies	South Rumaila
10.	Tewari and Dwivedi (2018b)	Ensemble methods	1 well data, 15 logs, 9 facies	Kansas, USA
11.	He et al. (2019)	Markov chain and ANNs	8 well data, 6 logs, 10 facies	Surat Basin, Australia
12.	Gu et al. (2019)	Probabilistic neural network, Boltzmann machine and particle swarm optimization	2 well data, 9 logs, 13 facies	IARA oilfield of Santos Basin
13.	Imamverdiyev and Sukhostat (2019)	Deep convolution neural network	10 wells data, 7 logs, 9 facies	Hugoton basin, Kansas, USA
14.	Bhattacharya et al. (2019)	SVM	554 well data, 5 logs, 7 facies	North Dakota, USA
15.	Moradi et al. (2019)	K-nearest neighbor, Naïve Bayes	4 well data, 7 logs, 6 facies	Upper Sarvak Formation, Iran

multidimensional and imbalanced petroleum data (Tewari and Dwivedi 2018b; Anifowose et al. 2015; Dietterich 2000; Skurichina and Duin 2001). However, ensemble methods are not properly explored and applied in the petroleum research domain. Only limited applications of the ensemble approach can be found in the petroleum domain as explained briefly in the next section of the paper.

The performance of HEMs, namely voting and stacked generalization ensembles, along with four other popular classifiers, has been studied for the recognition of mudstone lithofacies in the paper. This research work is primarily focused on the construction of diverse base classifiers contained in HEMs that can actually outperform single classifiers and HoEMs for lithofacies identification. A comparative study has been performed among HEMs and also with four contemporary classifiers for the prediction of lithofacies. The suitability of HEMs has also been evaluated for the identification of lithofacies. The complications associated with the development of HEMs have been discussed in this research work. Four popular supervised classifiers, viz. multilayer perceptron (MLP), SVM, gradient boosting (GB) and random forest (RF), have been combined in HEMs as base classifiers to provide more accurate and generalized results.

The performance of these classifiers has been evaluated using Kansas oil-field data with proper parameters optimization in their stable search ranges. Validation curve and grid search algorithm have been applied for parameters tuning to achieve maximum classification accuracy. The contribution of each input well logs, for the pattern recognition of mudstone lithofacies, has been studied using Relief algorithm. Relief algorithm is utilized for the attribute selections owing to its capability of identifying discriminatory information. Overall, this research work assesses the pattern recognition competency of HEMs for complex mudstone lithofacies.

Heterogeneous ensemble methods

Heterogeneous ensembles are lesser-known intelligent algorithms in the petroleum domain. They generate several different hypotheses in feature space using diverse base classifiers and combine them to achieve maximum possible accurate results. Sesmero et al. (2015) proved that diverse classification hypotheses in feature space are essential for the development of reliable, robust and generalized ensemble classifiers. Thus, HEMs are investigated in this paper

for solving multiclass lithofacies recognition problem in the quest for higher classification accuracy. Extraction of valuable lithofacies information from well logs data is quite a challenging task even for intelligent HoEMs. Spatial distribution and heterogeneous behavior of the hydrocarbon reservoir properties contribute to complexity, nonlinearly and uncertainty in all types of sensor-based measurements (Chaki et al. 2015; Bhattacharya et al. 2016). Also, no standard tools or techniques are available in the present scenario that can measure the reservoir heterogeneity and its influence on other reservoir properties, well logs and drilling data, etc. Therefore, HEMs are employed for recognition of subsurface lithofacies in this paper.

Related works

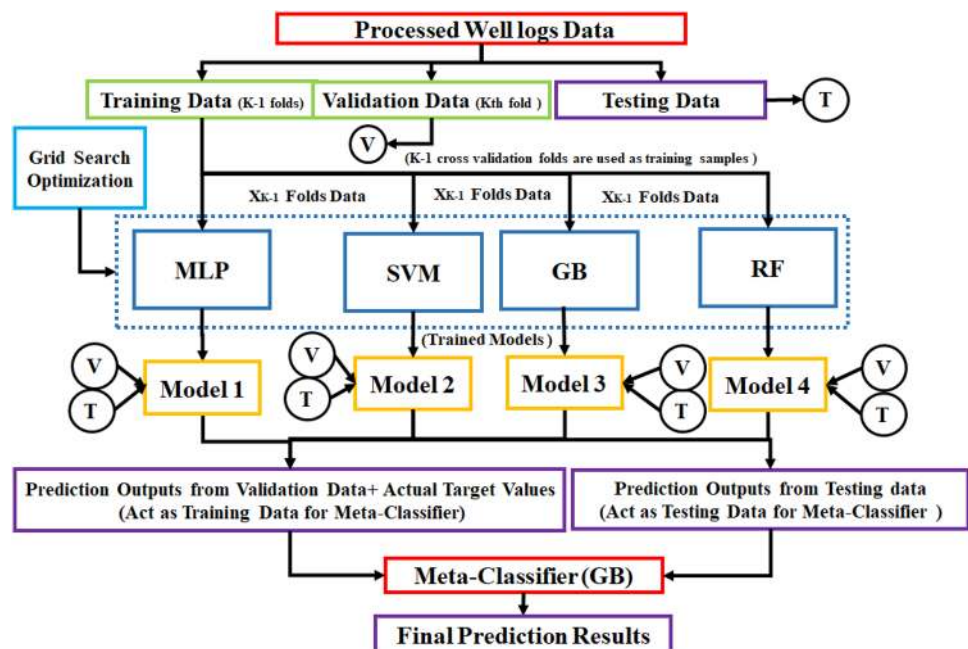
Ensemble methods are less applied methodology in the oil and gas industry. In the petroleum domain, HoEMs are mostly implemented for drilling parameter estimation and reservoir characterization. Only a few applications of HEMs are found in the existing literature. Santos et al. (2003) utilized a neural net ensemble for the recognition of underlying lithofacies. Gifford and Agad (2010) applied collaborative multiagent classification techniques for the identification of lithofacies. Masoudi et al. (2012) integrated the outputs of Bayesian and fuzzy classifiers to recognize productive zones in Sarvak Formation. Anifowose et al. (2015) utilized the stacked generalization ensemble for enhancing the prediction capability of supervised learners for reservoir characterization. Anifowose et al. (2017) wrote a review on the applications of ensemble methods and suggested that ensemble methods are suitable for solving problems of the oil and

gas industry. Bestagini et al. (2017) implemented the random forest ensemble for lithofacies classification of Kansas oil-field data. Xie et al. (2018) published a comparative study on the performance of HoEMs for the recognition of lithofacies. Tewari and Dwivedi (2018b) compared five HoEMs for the identification of lithofacies. Bhattacharya et al. (2019) applied ANN and random forest algorithms to predict daily gas production from the unconventional reservoir. Tewari et al. (2019) applied HoEMs for the prediction of reservoir recovery factors. HEMs have given higher classification performance as compared to supervised classifiers as well as HoEMs in several engineering fields such as remote sensing (Healey et al. 2018), prostate cancer detection (Wang et al. 2019), load forecasting (Ribeiro et al. 2019), wind speed forecasting (Liu and Chen 2019), etc. Therefore, HEMs are investigated in this paper to identify subsurface lithofacies. Two types of HEMs utilized for the identification of lithofacies are briefly explained below.

Stacked generalization ensemble

Stacked generalization ensemble is popularly known as stacking (Wolpert 1992). It combines decisions of different base classifiers in a single-ensemble architecture. Different base classifiers search the feature space with their diverse perspectives to find the maximum possible accurate hypotheses for a given classification task (Anifowose et al. 2015). The classification outcomes of base classifiers are combined together by a meta-classifier to provide final classification result. The combination of base classifiers' outcomes is decided by the meta-classifier algorithm. Figure 1 shows

Fig. 1 Conceptual architecture of the stacking ensemble utilized for quantitative lithofacies modeling in this study



the conceptual architecture of the stacking ensemble utilized for quantitative lithofacies modeling in this study.

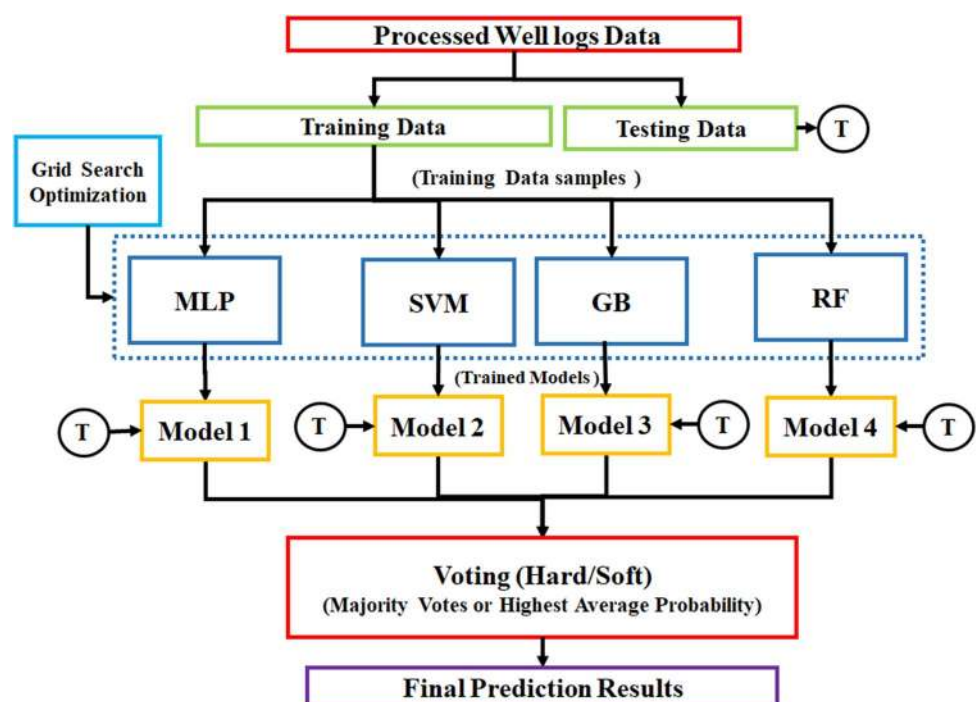
Stacking ensemble can also be created by merging the decision of similar base classifiers having different parametric values. The selection of base and meta-classifier combination is always a matter of concern during the design of stacking ensemble architecture. It is also difficult to design the most suitable configuration of classifiers in large feature space. Wolpert (1992) proved that the stacking ensemble is good in reducing the generalization error by decreasing bias and variance error associated with data. Initially, input data are split into training and testing datasets. Further, the training dataset is again split into K identical subsets similar to K -fold cross-validation technique. Base classifiers are trained on $(K - 1)$ subsets, while the K th subset is retained as a validation set. After training with $(K - 1)$ subsets, base classifiers are individually tested with the K th validation subset and also with the testing data. The outcomes of each base classifier with validation and test datasets will act as new training and testing data for meta-classifier. Moreover, the meta-classifier will be trained with the prediction outcomes of the validation set and the actual values of the target variable.

Voting ensemble

Voting ensemble combines the decisions of different base classifiers for the given classification or estimation task. It provides flexibility in combination strategies so

that the maximum possible classification accuracy can be achieved. It does not utilize any algorithm for the combination of predictions from base classifiers as in the stacking ensemble. Two combination schemes can be implemented for merging the decisions of several classifiers, namely majority vote rule (hard voting) and average predicted confidence probabilities (soft voting) to predict the class labels of test samples (Kittler et al. 1998). In place of meta-classifier, the abovementioned combining strategies are utilized to combine outcomes of diverse supervised classifiers. In hard voting, class labels of test samples are decided by majority voting rule. Every base classifier individually assigns a class label to a given test sample during the testing phase. The final classification of the test sample is decided by the maximum number of times a particular class label gets assigned to that test sample. On the other hand, soft voting strategy initially assigns weights to each base classifier. During the testing phase, it generates prediction probabilities for every test sample belonging to various classes. Later, these probabilities are multiplied with the weights assigned to every class labels and then it is averaged. Test samples are finally classified into that class which achieves the highest average confidence probability. Mathematically, soft voting technique classifies data samples as argmax (argument of maxima) of the sum of assigned probabilities (Kittler et al. 1998; Kuncheva 2004). Figure 2 shows theoretical architecture of voting ensemble utilized for quantitative lithofacies modeling in this study.

Fig. 2 Conceptual architecture of voting ensemble utilized for quantitative lithofacies modeling in this study



Experimental evaluation

In this paper, two HEMs were utilized for the recognition of quantitative lithofacies modeling. The primary goal of

this research work was to achieve higher classification performance using the HEMs approach. HEMs were trained and tested on real-field well logs data with other popular classifiers, namely RF, MLP, SVM and GB. All the ensemble methods were implemented on the Python Scikit-learn

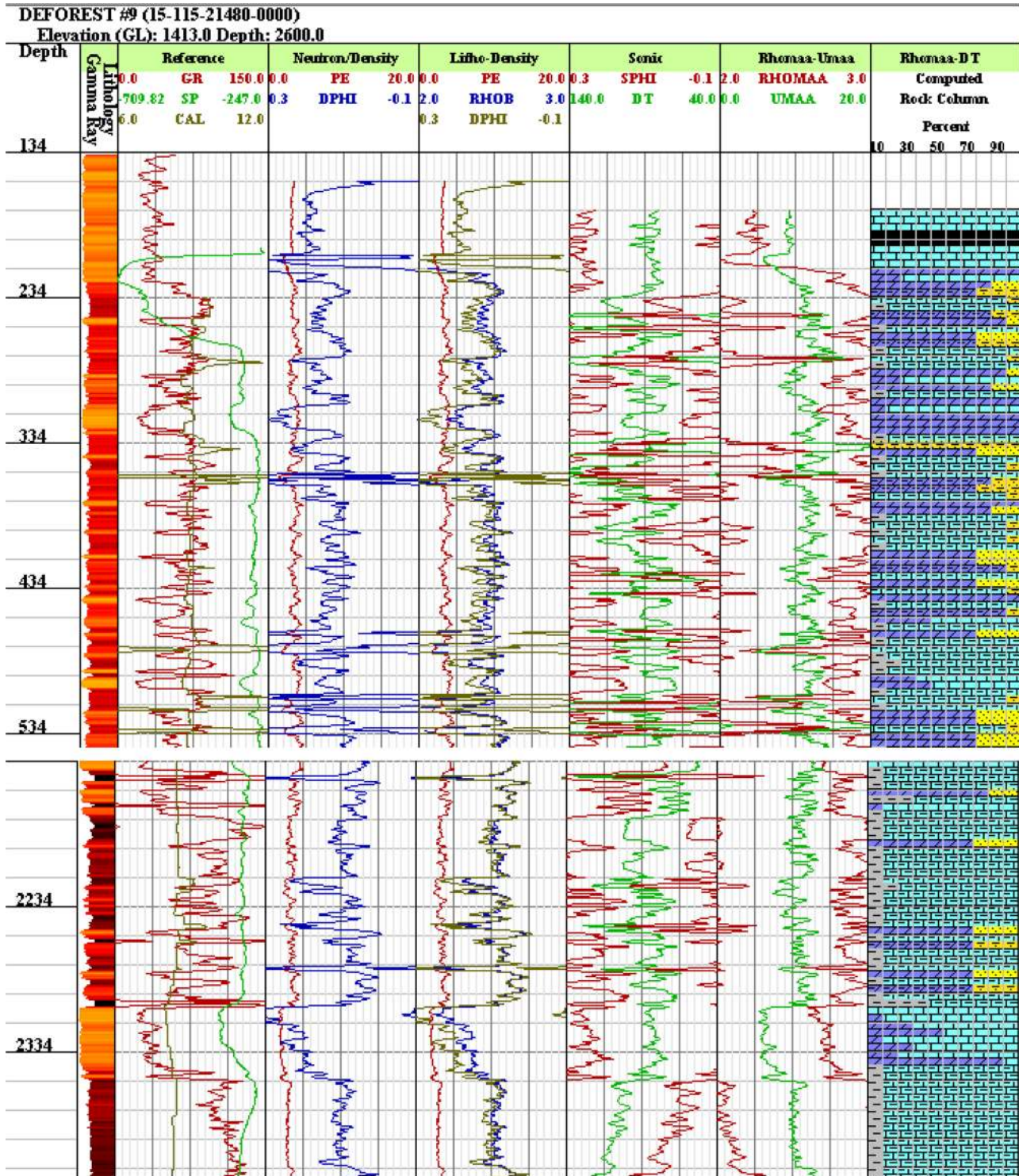


Fig. 3 Well logs of Kansas hydrocarbon-producing Deforest well (KGS database)

platform. Figure 3 portrays a portion of the geological lithofacies setting of Deforest well existing in the Kansas oil and gas field.

A brief description of Kansas field

Kansas region is mainly composed of sedimentary rocks with a maximum width of 2850 m. A large number of unconformities occur in the Kansas region with the sedimentary strata having 15–50% of the post-Precambrian period (Merriam 1963). Northeastern Kansas is enclosed by Pleistocene glacial deposits. A thick layer of Mesozoic rock is present in the western Kansas region. Mesozoic rock layers are mainly made up of limestones, chalks, sandstone, marine shales and nonmarine shale contents. Panoma field and Hugoton field, existing in western Kansas, comprise of large natural gas-producing reservoirs. Pennsylvanian and Permian systems are broadest structures of rock containing bedded rock salts in several layers. The pre-Pennsylvanian system existing in Kansas contains dolomites, marine, limestones layered alternatively with sandstones and shales. The Precambrian basement composed mainly of quartzite, granite and schist. Permian strata contain the carbonate reservoirs that produce the majority of natural gas. In 1992, Mississippian strata produced 43% of cumulative hydrocarbon production of Kansas field out of which 19% contributed to cumulative oil production (Newell 1987a). The numerous unconformities available in the Kansas region help trapping and migration of petroleum. Basal Pennsylvanian in Kansas has a huge deposition of hydrocarbon along with its length. A detailed description of the petroleum geology of the Kansas region can be found in Newell (1987a, b), Merriam (1963), Adler et al. (1971) and Jewett and Merriam (1959). Manual

interpretation of wells logs data of such a huge hydrocarbon-producing region is time-consuming and costly. Therefore, automatic detection and identification of subsurface lithofacies using machine learning algorithms is highly desirable to minimize cost and time.

Data description

The well logs data used for the development of ensemble methods were obtained from the Kansas Geological Survey (KGS) Web site (Kansas 2009) which is a very large available well logs data repository. The downloaded digital well logs, “Las” files, contain 13,000 data samples out of which 3425 samples are extracted belonging to nine different lithofacies, viz. dolomitic wackstone (DW) (1015), clay (CL) (320), dolomitic mudstone (DM) (240), dolomitic sandstone (DS) (455), siltstone (SS) (85), dolomitic packstone (DP) (265), carbonate mudstone (CM) (520), packstone (PS) (465) and wackstone (WS) (60). The above-said lithofacies are acknowledged as class labels for the classification of well logs data into their respective lithofacies. The downloaded “Las” files belong to Paradise A, Deforest and Strahm wells existing in the Kansas field. These files also contain information about mineral contents and lithofacies prevailing in these wells. “Las File Viewer” app has been downloaded from KGS Web site to visualize geological settings and facies of these wells. Table 2 contains the range of different well logs acting as input predictor variables. These logs and recorded information about different reservoir properties are used in the pattern recognition of lithofacies. The input well logs data were downloaded from the Kansas Geological Survey database (KGS) available on the KGS

Table 2 The statistical description of input well logs data

Well logs/units	Minimum	Maximum	Reservoir properties
Depth (DT)/ft	434	3771	Depth
Sonic log (SPOR)/pu	2.774	133.212	Formation porosity
Density log (DPOR)/pu	0.652	32.118	Formation density
Gamma ray log (GR)/API	16.781	414.152	Gamma radiation
Neutron log (NPOR)/pu	0	44.143	Formation porosity
Density correction log (RHOC)/(gm/cc)	0.01	0.296	Formation bulk density
Deep induction log (RILD)/ Ω m	1.905	65.158	Formation fluid saturation
Medium induction log (RILM)/ Ω m	2.031	140.706	Formation fluid saturation
Deep laterolog resistivity log (RLL3)/ Ω m	2.866	266.481	Formation fluid saturation
Caliper 1 log (DCAL)/inch	7.468	8.982	Borehole diameter
Caliper 2 log (MCAL)/inch	6.947	8.911	Borehole diameter
Microinverse resistivity log (MI)/ Ω m	0.609	41.818	Flushed zone resistivity
Micronormal resistivity log (MN)/ Ω m	0.138	44.097	Flushed zone resistivity
Acoustic transit time 1 log (DT)/(U s/ft)	51.522	235.961	Compaction, stratigraphic, etc.
Acoustic transit time 2 log (TT1)/(U s/ft)	54.36	662.885	Compaction, stratigraphic, etc.
Spontaneous potential log (SP)/MV	−1.797	71.61	Formation thickness and boundaries

Web site only for research purposes. The well logs are also accessible in the form of digital logs (.csv) files. The digital logs (.csv) files contain null or missing values in the data samples that are generated due to calibration loss of logging sensors, faulty logging tools, etc. Figure 4 shows generalized workflow for the HEMs to recognize the subsurface lithofacies.

Data preprocessing

Diaz et al. (2018) suggested that the preprocessing of petroleum data, such as resampling, normalization, noise filtering, attribute selection, etc., helps to improve the classification or estimation accuracy of intelligent algorithms. Initially, resampling of well logs data was done to eliminate samples containing null, garbage and missing values. After resampling, the input data were normalized to reduce the impact of larger values on the smaller values of predictor variables. Input data can be normalized as given below.

$$X_{i,\text{norm}} = \frac{X_i - X_{\text{Min}}}{X_{\text{Max}} - X_{\text{Min}}} \quad (1)$$

where X_{Max} and X_{Min} are maximum and minimum values of the respective predictor variables. Equation (1) represents the Min–Max normalization technique that ensures all the input variables are equally scaled. Min–Max normalization is preferred over other scaling techniques especially when input data distribution does not follow Gaussian or normal distribution. It is highly beneficial for those machine learning algorithms that involve distance calculation or optimization in their internal mathematical architectures such as ANNs, K-means clustering, SVM, etc. (Mustafa and Yusof 2010; Shalabi et al. 2006; Suarez-Alvarez et al. 2012).

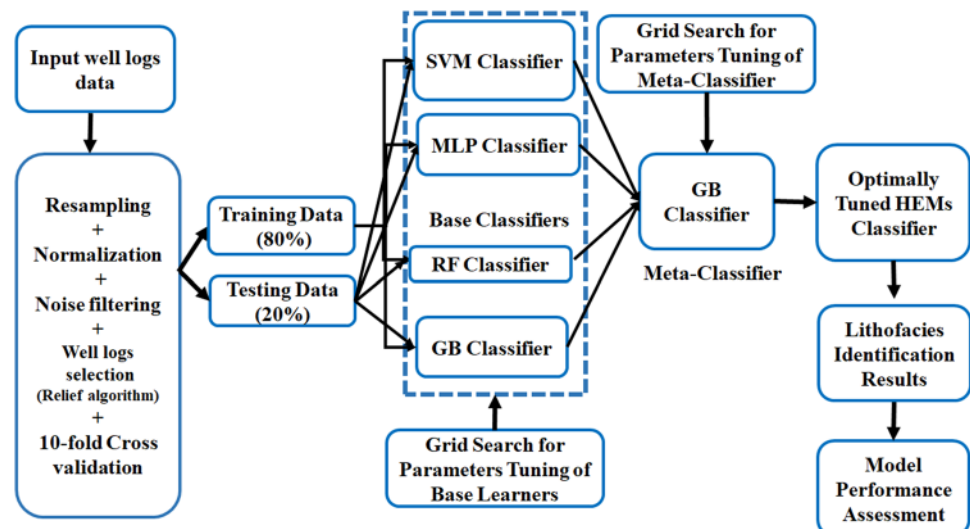
Noise filtering

Noise filtering of well logs data was done to minimize the effects of noise during the pattern recognition of lithofacies. Tewari and Dwivedi (2019) studied the influence of noise levels on the performance of supervised classifiers and reported its damaging effects on the classifier's performance. There are several denoising techniques that are available in the petroleum and geophysics literature such as low-pass filter, high-pass filter, Savitzky–Golay, wavelet denoising, moving average, Gaussian, etc. Savitzky–Golay (SG) smoothing filter has been found suitable and widely utilized noise filtering technique for geophysical data (Baba et al. 2014). This is a digital filter used for data smoothening which fits a polynomial of degree n by the linear least square method and maintains signal tendency through convolution (Baba et al. 2014). High peaks of well logs data were considered as noise components which were eliminated using SG filters. Noise contents harmfully affect the pattern recognition ability of intelligent classifiers. Figure 5 shows four important well logs with original and denoised waveform. SG smoothening filter was utilized for removing the noise content of input well logs. The degree of polynomial fitted in well logs for smoothening was found to be varying from 5 to 13. The higher components or spikes in well logs were treated as noisy contents and eliminated through smoothening the waveforms of well logs. Fifteen input well logs data were smoothened using the SG filter and passed to Relief algorithm for important well logs selection.

Attribute selection

Important well logs were selected to decrease the dimensionality of data by removing the redundant logs. The high dimensionality of logs data increases computational cost and

Fig. 4 A generalized conceptual workflow for the HEMs to recognize the subsurface lithofacies



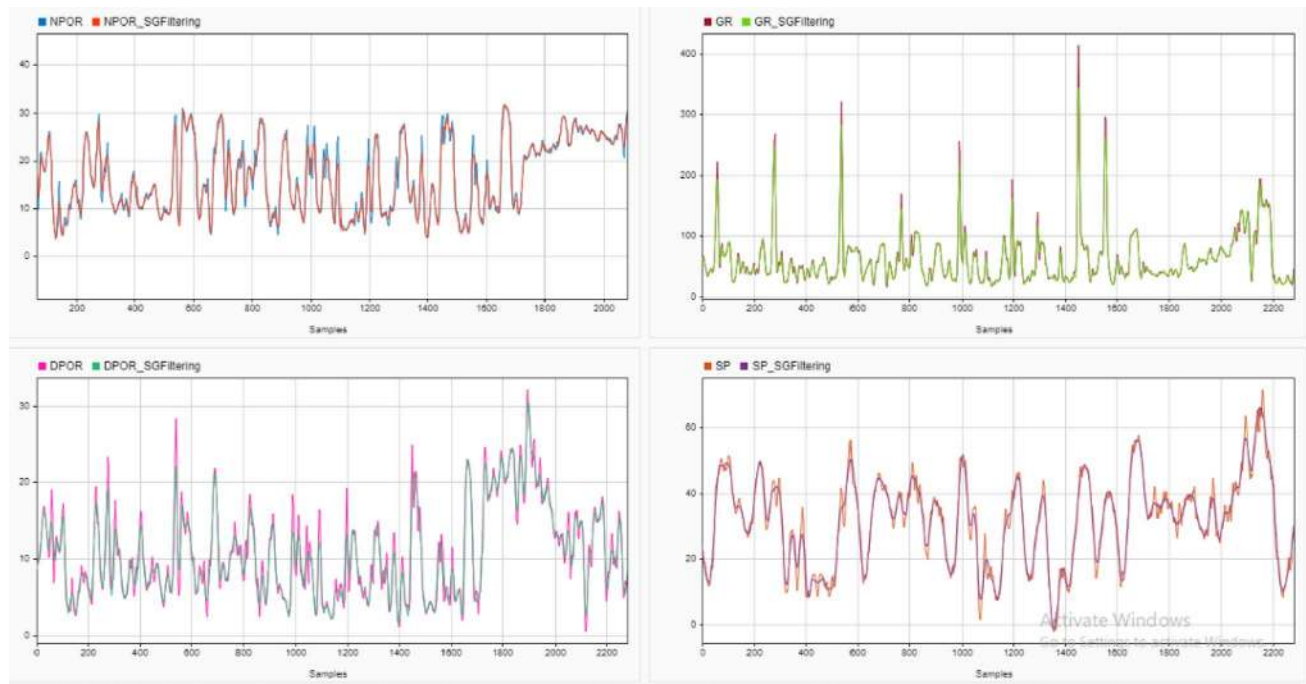


Fig. 5 Four original well logs showed with their denoised waveforms using SG smoothing filter

time during the pattern recognition of lithofacies. Several attribute selection paradigms are available in the literature such as a forest of tree-based attributes selection, univariate feature selection, Relief algorithm, etc. Overlapping lithofacies can only be identified by recognizing their discriminatory information contained in the attributes or logs. Discriminatory information plays a decisive role, especially in classification tasks. Further, the Relief algorithm was applied for the attributes selection due to its capability of identifying discriminatory information. Relief algorithm recognizes conditional dependencies and correlations among the attributes or predictor variables and lithofacies (Jia et al. 2013; Farshad and Sadeh 2014; Urbanowicz et al. 2017). Relief algorithm assigns weights and ranks to every predictor variable depending upon their relevance for the pattern recognition of lithofacies. Figure 6 shows input well logs with predictor importance weights plotted on the y -axis and ranks on the x -axis. The well logs having negative weights were removed as they did not add any contribution in the pattern recognition process. NPOR, GR, DPOR, SP, MI, MN, SPOR, DT and RILD were important well logs contributing to the identification and recognition of mudstone lithofacies as shown in Fig. 6.

Data partition

The processed input data were further divided into training sets and testing sets using a cross-validation technique. There are three cross-validation techniques, namely K-fold,

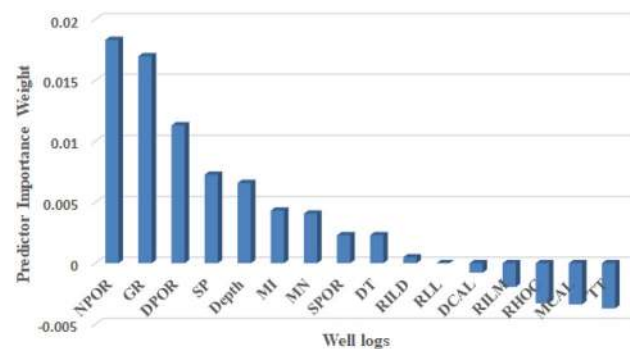


Fig. 6 Available well logs arranged according to their predictor important weights assigned by Relief algorithm for pattern recognition of lithofacies

leave-one-out and hold out, that are popular in the machine learning domain for the generation of training and testing datasets. K-fold cross-validation technique was utilized in this research work for splitting the processed input data into training and testing subsets ($K = 10$). Tenfold cross-validation (10-FCV) technique has been reported to have minimum variance error as compared to other cross-validation techniques (Kohavi 1995). Cross-validation helps to minimize the chances of overfitting and underfitting of models (Xie et al. 2018). The input well logs data were randomly divided into K subsets during K-FCV. $(K - 1)$ subsets were used for training the intelligent models and K th for testing it. This was repeated in iterations until each subset gets selected as

a testing set. The final accuracy of every machine learning model was decided by averaging the accuracies obtained in the iterations.

Optimization of model parameters

The optimum value of every model parameter is essential to be determined during the training phase so that models can be generalized for unseen data samples. Grid search algorithm was utilized for tuning model parameters to achieve maximum classification accuracy on unseen test data samples. Models with optimally tuned parameters were saved, to classify unseen test data samples, and were also evaluated for their generalizability and reliability. Grid search algorithm is one of the popular tuning paradigms in the petroleum domain (Tewari et al. 2019a). Machine learning models always have the possibility of getting overfitted or underfitted during pattern recognition. A separate validation score test was conducted to examine the overfitting and underfitting tendency of intelligent models. Validation curve was utilized to shrink the search range for various parameters. It clearly illustrates the overfitting and underfitting regions of the respective classifiers with a specific parameter variation. In an underfitting state of the intelligent model,

training and validation scores are normally recorded to be low, whereas overfitting states result in high training and low validation scores. The parameter search range is primarily comprised of upper and lower constraints of a stable region. In stable region, no dramatic variation in training and validation scores takes place. However, the model still needs an optimization algorithm that explores within the stable search range to find the best possible value of the model parameters. The search range and optimum values for various model parameters are depicted in Table 3. Figures 7 and 8 show the validation curves of GB and RF classifiers for four important parameters, namely Estimators, Min_samples_split, Max_depth and Min_samples_leaf. Figure 9a, b shows the validation curves of SVM for regularization constant (C) and gamma (γ) versus accuracy score (Fig. 9).

The optimal settings of MLP parameters were determined through several computational trials to obtain the maximum possible classification accuracy. The speed of convergence and training loss function of MLP decides its classification performance during training and testing phases. Figure 10 compares several learning strategies available for training of MLP classifier using training loss versus iterations plots. The initial learning rate of MLP was set at 0.001 for lithofacies prediction. MLP utilized Adam solver for weight

Table 3 The search range and optimized values of model parameters obtained through grid search algorithm on input well logs data

S. no.	Classifiers	Model parameters	Search range	Settings
1.	MLP	Activation function	Identity, logistic, tanh, relu	relu
		Solver	Lbfgs, Sgd, Adam	Adam
		alpha	0.00001–0.1	0.0001
		Learning_rate_init	0.0001–0.1	0.001
		Learning_rate	Constant, invscaling and adaptive	adaptive
		Max_iteration	10–400	200
2.	SVM	C	1–1000	200
		gamma	0.001–1	4
		Kernel	RBF, polynomial, linear	RBF
3.	RF	Estimators	10–1000	100
		Max_depth	0-infinity	10
		Min_samples_split	2–10	2
		Min_samples_leaf	1–10	1
4.	GB	Estimators	10–400	250
		Max_depth	0-infinity	None
		Min_samples_split	2–10	2
		Min_samples_leaf	1–10	1
5.	Voting	Base classifiers	Any supervised classifiers and HoEMs	RF, GB, MLP, SVM
		Weights	0.1–1	1,1,0.5,0.5,1
		Voting	Soft/hard	Hard
		Combining strategies	Simple majority voting/weighted majority voting rule	Weighted majority voting rule
6.	Stacking	Base classifiers	Any supervised classifiers and HoEMs	RF, GB, MLP and SVM
		Meta-classifier	RF, GB, MLP and SVM	GB

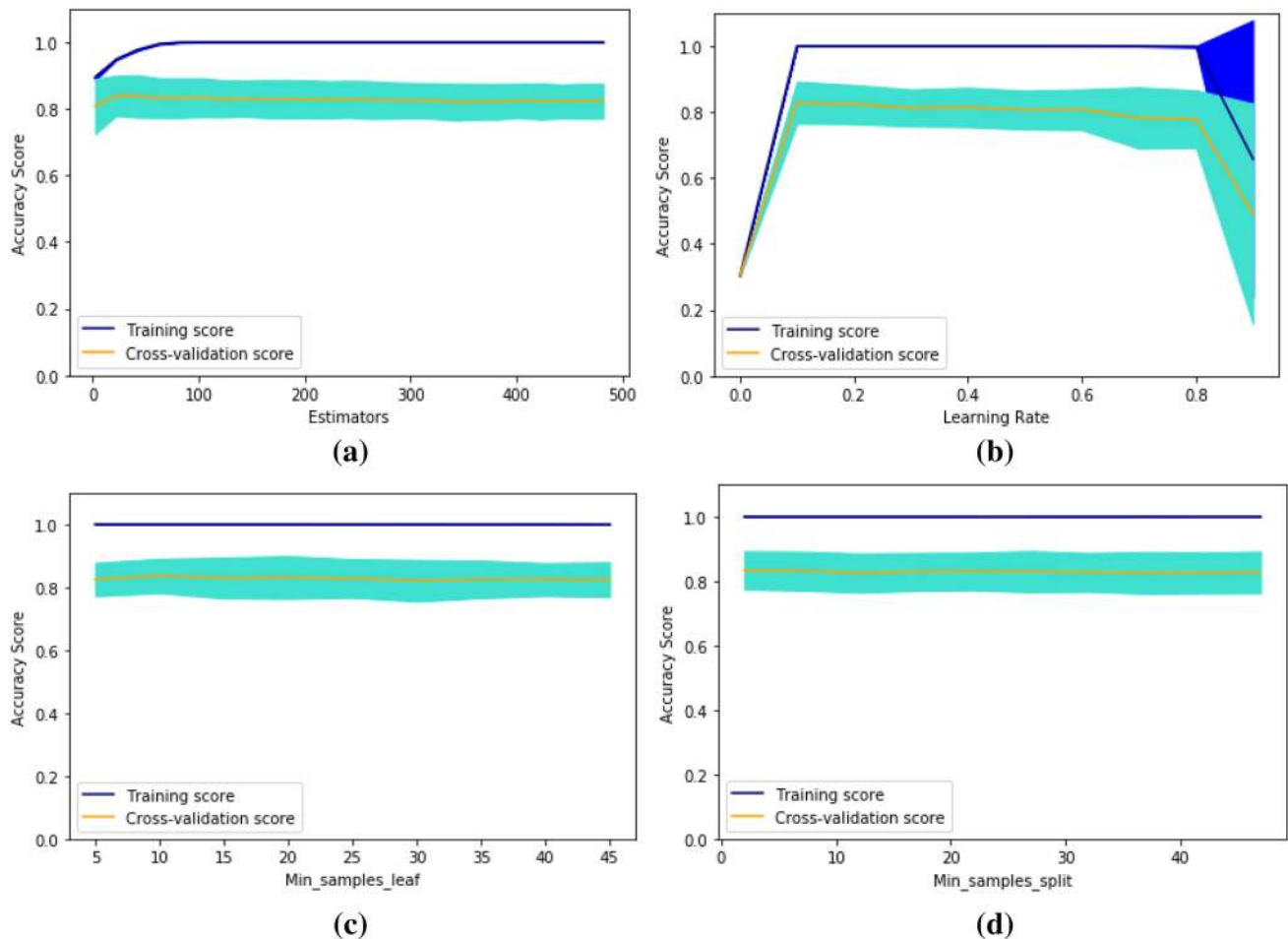


Fig. 7 Validation curves of gradient boosting classifier to identify stable search range for four primary model variables. **a** Number of estimators, **b** learning rate, **c** minimum samples required at a leaf node and **(d)** minimum samples required for splitting the internal node

optimization because of its fast convergence and low training loss for large data as depicted in Fig. 10. MLP updates its parameters iteratively during training operation using the fractional derivatives of the loss function. Cross-entropy loss function was utilized to calculate the probability for data samples belonging to a particular class or lithofacies. Figure 10 shows the comparative graph for diverse training strategies of MLP.

Performance evaluation

The performance of optimally tuned intelligent models was evaluated with testing data using five statistical performance indicators, namely recall, precision, *f1*-score, accuracy and Matthew correlation coefficients (MCC). The accuracy parameter was used to evaluate the classification performance of each classifier for the recognition of lithofacies. This parameter is recommended only for balance data conditions and becomes unreliable with uneven data distribution. Precision and recall also act as performance metrics for classifiers. Every classifier

should maximize the values of precision and recall for good classification results. *F1*-score investigates the accuracy of precision and recall value and mostly used in information retrieval domains. However, in the case of data with high imbalance conditions, these performance indicators may give misleading results. Therefore, the test performance of each classification model is also evaluated using the MCC parameter which is unaffected by data imbalance issues. The performance indicators used for the evaluation of machine learning models are given below.

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (3)$$

$$f1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

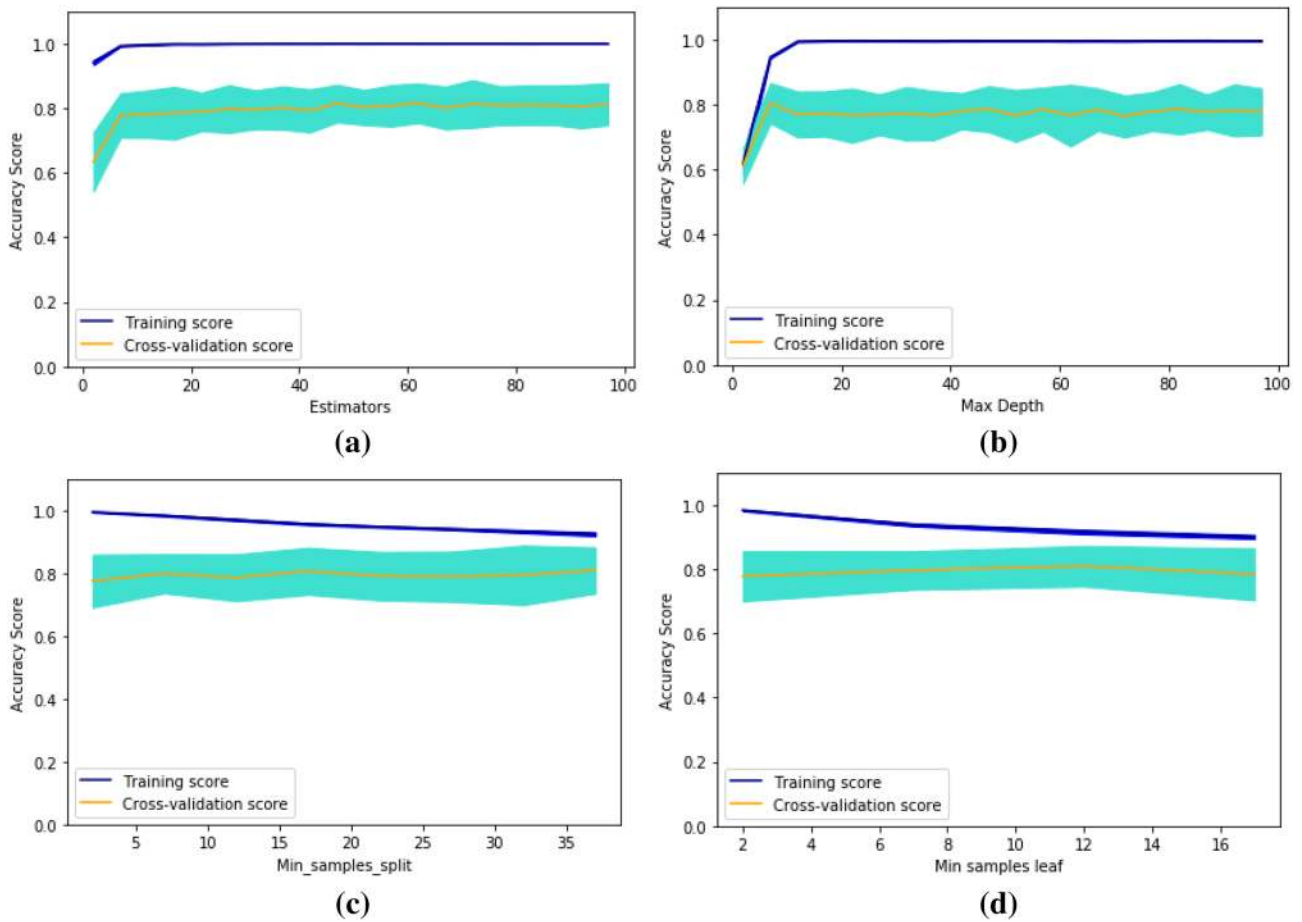


Fig. 8 Validation curves for RF classifier to identify stable search range for four primary model parameters. **a** Number of estimators, **b** maximum depth of tree, **c** minimum samples required for splitting the internal node, **d** Minimum samples required at the leaf node

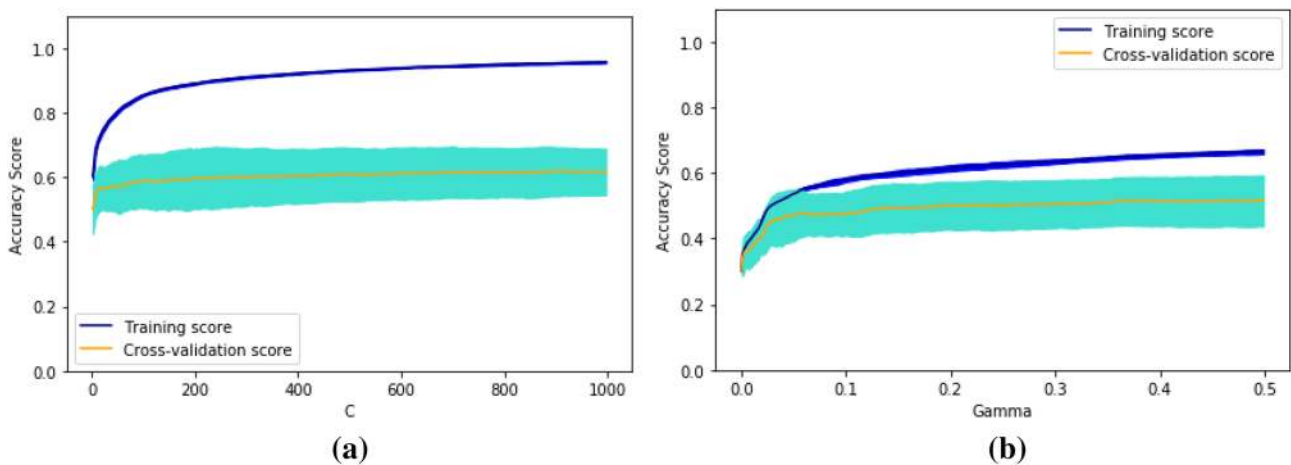


Fig. 9 Validation curves for SVM classifier to identify stable search range for two primary model parameters. **a** Penalty cost parameters for misclassified error samples and **b** kernel coefficient of RBF

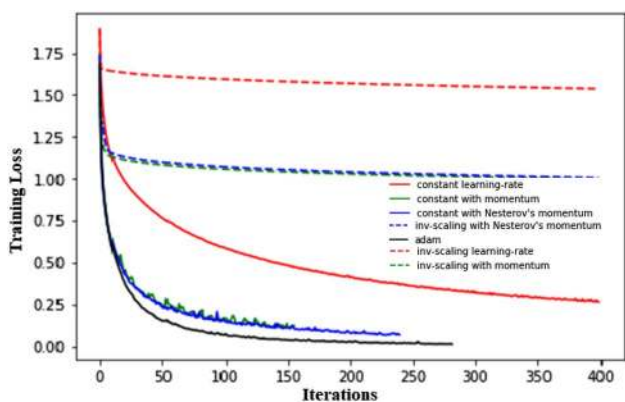


Fig. 10 Comparison among diverse learning strategies for MLP classifier for prediction of lithofacies

$$\text{Accuracy} = \frac{\text{Correctly classified data samples}}{\text{Total number of data samples}} \tag{5}$$

$$\text{MCC} = \frac{\text{SC} \times \text{TS} - \sum_k \text{PC}_k \times \text{TC}_k}{\sqrt{(\text{TS}^2 - \sum_k \text{PC}_k^2) \times (\text{TS}^2 - \sum_k \text{TC}_k^2)}} \tag{6}$$

where TP is a number of correctly classified data samples of target lithofacies, FP is a number of correctly classified data samples other than target lithofacies, and FN is the number of incorrectly recognized samples classified as target lithofacies. In MCC, TC_k is the number of times prediction of k class truly happened, SC is correctly classified data samples, TS is the total number of data samples, and PC_k is the number of times k class predicted.

Results and discussion

This section discusses the experimental results obtained during the recognition of nine mudstone lithofacies belonging to Kansas oil and gas fields. The performance of stacking and voting ensembles was compared with four popular classifiers, namely GB (Friedman 2001), RF (Ho 1995), SVM (Cortes and Vapnik 1995) and MLP (Windeatt 2006). Stacking and voting are two HEMs that were implemented to predict the complex lithofacies. Figure 4 depicts a generalized conceptual workflow for HEMs to predict lithofacies of the formations. The performance of HEMs was tested by two separate data-driven experiments for the prediction of lithofacies. In the first experiment, 10-FCV was performed to split the input data samples into training and testing subsets so that generalized prediction outcomes can be obtained. The performance of each classifier has been reported in the form of precision, recall and $f1$ -score for individual lithofacies. Tables 4, 5 and 6 show precision, recall and $f1$ -score acquired by HEMs and base classifiers for each lithofacies during 10-FCV. Overall, the classification performance of stacking has been found higher than all the other classifiers considered in this study. Voting ensemble has secured second place in terms of overall classification performance as shown in Tables 4, 5 and 6. GB and RF classifiers have given similar performance scores for the identification of mudstone lithofacies as shown in Table 5. SVM classifier has also maintained good classification performance during 10-FCV for all the lithofacies. MLP becomes the worst performing classifier in terms of evaluation metrics, viz. average precision, average recall and average $f1$ -score, as shown in Table 6. It is also found that voting, GB, RF and MLP have fluctuations in their performances for smaller classes, namely SS and WS. However, stacking and SVM classifiers are successful in maintaining their performances even for smaller classes as shown in Tables 4 and 6. Smaller classes

Table 4 The performance of HEMs after tenfold cross-validation for lithofacies classification

Stacking ensemble				Voting ensemble			
Lithofacies	Precision	Recall	$f1$ -score	Lithofacies	Precision	Recall	$f1$ -score
CL	0.77	0.86	0.81	CL	0.80	0.80	0.80
CM	0.78	0.89	0.83	CM	0.76	0.87	0.81
DM	0.92	0.96	0.94	DM	0.98	0.88	0.92
DP	0.80	0.68	0.73	DP	0.86	0.60	0.71
DS	0.87	0.85	0.86	DS	0.74	0.92	0.82
DW	0.96	0.93	0.95	DW	0.95	0.94	0.95
PS	0.93	0.83	0.87	PS	0.94	0.83	0.88
SS	0.76	0.94	0.84	SS	0.8	0.94	0.86
WS	1.00	0.83	0.91	WS	1.00	0.58	0.74
Microavg.	0.87	0.87	0.87	Microavg.	0.86	0.86	0.86
Macroavg.	0.87	0.86	0.86	Macroavg.	0.87	0.82	0.83
Weighted avg.	0.88	0.87	0.87	Weighted avg.	0.87	0.86	0.86

have contributed a lesser number of data samples during training and testing of machine learning models. These classes also represent facies having thin layers that are difficult to identify using conventional well logs interpretation techniques. WS and SS facies are intentionally included with lesser data samples to magnify data imbalance conditions that make classification harder even for strong classifiers

such as GB, RF, voting, etc. Voting and stacking ensembles have utilized the same base classifiers for the classification of facies; however, stacking performed better than voting due to the presence of meta-classifier for combining the outcomes of base classifiers.

In the second experiment, a separate test was also performed with randomly selected training and testing data

Table 5 The performance of GB classifier and RF ensembles after tenfold cross-validation for lithofacies classification

GB classifier				RF classifier			
Lithofacies	Precision	Recall	<i>f1</i> -score	Lithofacies	Precision	Recall	<i>f1</i> -score
CL	0.81	0.75	0.78	CL	0.79	0.75	0.77
CM	0.74	0.87	0.8	CM	0.76	0.87	0.81
DM	0.94	0.94	0.94	DM	0.98	0.85	0.91
DP	0.85	0.64	0.73	DP	0.87	0.62	0.73
DS	0.75	0.89	0.81	DS	0.73	0.9	0.81
DW	0.93	0.95	0.94	DW	0.94	0.95	0.94
PS	0.95	0.78	0.86	PS	0.89	0.80	0.84
SS	0.78	0.82	0.8	WS	0.8	0.94	0.86
WS	0.88	0.58	0.7	SS	0.83	0.42	0.56
Microavg.	0.85	0.85	0.85	Microavg.	0.85	0.85	0.85
Macroavg.	0.85	0.8	0.82	Macroavg.	0.84	0.79	0.8
Weighted avg.	0.86	0.85	0.85	Weighted avg.	0.86	0.85	0.85

Table 6 The performance of SVM and MLP classifiers after tenfold cross-validation for lithofacies classification

SVM classifier				MLP classifier			
Lithofacies	Precision	Recall	<i>f1</i> -score	Lithofacies	Precision	Recall	<i>f1</i> -score
CL	0.79	0.81	0.80	CL	0.38	0.33	0.35
CM	0.80	0.90	0.85	CM	0.57	0.78	0.66
DM	0.94	0.65	0.77	DM	0.70	0.67	0.68
DP	0.83	0.74	0.78	DP	0.56	0.19	0.28
DS	0.70	0.89	0.79	DS	0.52	0.73	0.6
DW	0.97	0.94	0.95	DW	0.94	0.89	0.91
PS	0.97	0.84	0.90	PS	0.78	0.78	0.78
SS	0.8	0.94	0.86	SS	0.67	0.47	0.55
WS	0.98	0.92	0.96	WS	0.0	0.0	0.0
Microavg.	0.86	0.86	0.86	Microavg.	0.69	0.69	0.69
Macroavg.	0.87	0.85	0.85	Macroavg.	0.57	0.54	0.54
Weighted avg.	0.86	0.86	0.87	Weighted avg.	0.68	0.69	0.67

Table 7 Classification accuracy of six machine learning models depicted on training (80%) and testing (20%) datasets for lithofacies classification

S. no.	Classifier	Base classifier	Training accuracy	Testing accuracy	Precision	Recall	<i>f1</i> -score	MCC
1.	MLP		0.6892	0.6554	0.67	0.67	0.63	0.63
2.	SVM		0.9170	0.8403	0.86	0.87	0.87	0.81
3.	GB	Decision tree	0.9105	0.8554	0.86	0.86	0.86	0.83
4.	RF	Decision tree	0.9012	0.8481	0.86	0.85	0.85	0.83
5.	Voting	MLP, SVM, RF, GB	0.9230	0.8657	0.87	0.87	0.87	0.85
6.	Stacking	MLP, SVM, RF, GB	0.9278	0.8832	0.89	0.88	0.88	0.86

Fig. 11 Confusion matrices for different classifiers using 20% of input data as a testing dataset to predict different lithofacies

		Predicted Lithofacies								
		CL	CM	DM	DP	DS	DW	PS	SS	WS
	CL	55/64	4	0	0	2	3	0	0	0
	CM	2	93/104	0	4	1	4	0	0	0
	DM	0	0	46/48	0	2	0	0	0	0
	DP	0	6	2	36/53	5	0	3	1	0
	DS	6	2	2	4	77/91	0	0	0	0
	DW	3	10	0	0	0	189/203	1	0	0
	PS	5	5	0	0	2	0	77/93	4	0
	SS	0	0	0	1	0	0	0	16/17	0
	WS	0	0	0	0	0	0	2	0	10/12
	Accuracy	85.9%	89.4%	95.8%	67.9%	84.6%	93.1%	82.8%	94.1%	83.3%

(a) Confusion matrix for Stacking ensemble.

		Predicted Lithofacies								
		CL	CM	DM	DP	DS	DW	PS	SS	WS
	CL	51/64	4	0	0	6	3	0	0	0
	CM	2	90/104	0	2	4	6	0	0	0
	DM	0	0	42/48	0	6	0	0	0	0
	DP	1	7	1	32/53	10	0	1	1	0
	DS	5	1	0	1	84/91	0	0	0	0
	DW	3	9	0	0	0	191/203	0	0	0
	PS	2	8	0	0	3	0	77/93	3	0
	SS	0	0	0	1	0	0	0	16/17	0
	WS	0	0	0	1	0	0	4	0	7/12
	Accuracy	79.7%	86.5%	87.5%	60.4%	92.3%	94.1%	82.8%	94.1%	58.3%

(b) Confusion matrix for Voting ensemble.

		Predicted Lithofacies								
		CL	CM	DM	DP	DS	DW	PS	SS	WS
	CL	48/64	3	0	0	7	6	0	0	0
	CM	1	91/104	0	1	5	6	0	0	0
	DM	0	0	45/48	0	3	0	0	0	0
	DP	0	9	1	34/53	9	0	0	0	0
	DS	5	1	1	3	81/91	0	0	0	0
	DW	2	10	0	0	0	191/203	0	0	0
	PS	3	9	0	0	2	1	73/93	4	1
	SS	0	0	1	1	1	0	0	14/17	0
	WS	0	0	0	1	0	0	4	0	7/12
	Accuracy	75%	87.5%	93.8%	64.2%	89%	94.1%	78.5%	82.4%	58.3%

(c) Confusion matrix for GB classifier.

samples without 10-FCV. Table 7 depicts the overall performance of every classifier utilized in this study with processed input data split into (80%) training subset and (20%)

testing subset. The testing accuracy for individual lithofacies is depicted diagonally in confusion matrices. Figure 6 illustrates the confusion matrices of HEMs and their base

Fig. 11 (continued)

		Predicted Lithofacies(RF)								
		CL	CM	DM	DP	DS	DW	PS	SS	WS
CL	48/64	4	0	0	6	6	0	0	0	0
CM	1	90/104	0	2	5	6	0	0	0	0
DM	0	0	41/48	0	7	0	0	0	0	0
DP	1	6	1	33/53	9	0	3	0	0	0
DS	6	2	0	1	82/91	0	0	0	0	0
DW	2	9	0	0	0	192/203	0	0	0	0
PS	3	8	0	0	3	0	74/93	4	1	0
SS	0	0	0	1	0	0	0	16/17	0	0
WS	0	0	0	1	0	0	6	0	5/12	0
Accuracy	75%	86.5%	85.4%	62.3%	90.1%	94.6%	79.6%	94.1%	41.7%	

(d) Confusion matrix for RF classifier.

		Predicted Lithofacies								
		CL	CM	DM	DP	DS	DW	PS	SS	WS
CL	47/64	2	0	0	12	2	1	0	0	0
CM	2	94/104	0	3	0	4	1	0	0	0
DM	0	0	29/48	0	19	0	0	0	0	0
DP	2	6	0	36/53	5	0	3	1	0	0
DS	4	2	1	1	83/91	0	0	0	0	0
DW	4	8	0	0	0	191/203	0	0	0	0
PS	4	6	0	2	2	0	77/93	1	1	0
SS	0	0	0	2	0	0	0	15/17	0	0
WS	0	0	0	0	0	0	5	0	7/12	0
Accuracy	73.4%	90.4%	60.4%	67.9%	91.2%	94.1%	82.8%	88.2%	58.3%	

(e) Confusion matrix for SVM classifier.

		Predicted Lithofacies								
		CL	CM	DM	DP	DS	DW	PS	SS	WS
CL	21/64	10	0	0	24	2	7	0	0	0
CM	10	81/104	0	3	0	9	1	0	0	0
DM	1	0	32/48	0	15	0	0	0	0	0
DP	5	18	0	10/53	15	0	2	3	0	0
DS	7	2	14	2	66/91	0	0	0	0	0
DW	4	19	0	0	0	180/203	0	0	0	0
PS	6	9	0	0	4	0	73/93	1	0	0
SS	0	2	0	3	4	0	0	8/17	0	0
WS	2	0	0	0	0	0	10	0	0/12	0
Accuracy	32.8%	77.9%	66.7%	18.9%	72.5%	88.7%	78.5%	47.1%	0.0%	

(f) Confusion matrix for MLP classifier.

classifiers generated during the testing phase. Training and testing classification accuracies of HEMs are found higher than all other machine learning models utilized in this study as shown in Table 7. Naturally, subsurface layers exist inside the formations with uneven thickness and patterns. Therefore, uneven data distribution has been considered to represent real-field conditions. This also provides us an opportunity to understand worst to best possible performance of machine learning classifiers for individual layers during imbalance data conditions. The uneven data distribution is in particular chosen for this study to understand the effect of data imbalance conditions. Facies having lesser data points such as WS, SS, etc., are designed for magnifying data imbalance effects. Stacking ensemble has shown great potential to extract lithofacies information from well logs data even for smaller classes due to the presence of meta-classifier in its architecture. Stacking ensemble has scored 83% accuracy for WS and 94% for SS which are challenging smaller lithofacies. This research work is specially designed to evaluate worst- to best-case scenarios for lithofacies modeling. Layerwise classification accuracy of HEMs along with its base classifiers can be summarized as follows: (a) stacking (67.9–95.8%), (b) voting (58.3–94.1%), (c) GB (58.3–94.1%), (d) RF (41.7–94.6%), (e) SVM (58.3–94.1%) and (f) MLP (0.0–88.7%). In the case of data with high imbalance conditions, performance indicators (viz. accuracy, precision, recall, and *f1*-score) may give misleading results. Therefore, the testing performance of each classification model is also evaluated using the MCC parameter which is unaffected by data imbalance issues as shown in Table 7. It is found that MCC scores of applied models also justify their performance as shown in Table 7. DP has emerged as one of the most challenging subsurface rock layers during the testing phase. In this study, all the classifiers have identified data samples related to DP mostly as CM. It may be possible that the presence of calcareous mud inside DP has confused base classifiers with CM. This uncertainty may be removed by increasing the number of training data samples that will help in learning discriminating features between similar layers (Fig. 11).

Conclusions

A rigorous facieswise comparison has been made between stacking and voting ensembles for the detection and identification of lithofacies. Stacking has shown nearly 4% and 2%

improvement in test accuracy as compared to SVM and RF. Four popular machine learning algorithms have been combined in HEMs as base classifiers to provide more accurate and generalized results. In this study, HEMs have combined MLP, SVM, GB and RF classifiers to achieve better classification accuracy than their individual performances. The individual performance of the abovementioned classifiers has been evaluated using Kansas oil and field data with proper parameter optimization in their stable search ranges. Validation curve and grid search algorithm have been properly utilized for the model parameters tuning to achieve maximum classification accuracy. The research work carried out in this paper has led to the following conclusions.

- The performance of HEMs depends upon the selection of efficient base classifiers for the quantitative lithofacies modeling.
- Validation curve has been found as an efficient measure for identifying stable search range for machine learning parameters.
- Stacking ensemble has shown great potential to extract lithofacies information from well logs data.
- The training and testing classification accuracies of HEMs have been found highest among the other classifiers used in this study.
- DP layer is found to be the most challenging facies among all the nine target lithofacies. Stacking ensemble has given the highest individual identification accuracy for all the layers of lithofacies.
- Prediction accuracy of individual facies ranges from 67.9 to 95.8% (worst to best possible testing accuracy), and maximum overall accuracy is (training = 92.78% and testing = 88.32%) obtained for stacking ensemble.

In this study, HEMs have shown its potential for quantitative lithofacies modeling and have outperformed the other classifiers. A combination of diverse base classifiers will lead to higher accuracy and better model generalization as established from the results obtained in this study. The analysis of results reveals that HEMs are practical and more accurate models, with a significant improvement in classification accuracy for lithofacies identification, as compared to the individual base classifiers.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

Table 8 A brief description of various important models' parameters utilized in this study for HEMs

S. no.	Classifiers	Model parameters	Description of models' parameters
1.	MLP	Activation function	These are nonlinear differentiable mathematical functions that help MLP to correlate complex input variables with response or target variables. It is also essential for backpropagation optimization of weights and error reduction, e.g., “relu,” “logistic,” etc.
		Solver	It provides weights optimization strategy to MLP, e.g., “lbfgs,” “sgd,” etc.
		alpha	MLP is penalty parameter for misclassification.
		Learning_rate_init	It decides the step size of weights updating and provides initial value to learning rate
		Learning_rate	It decides rate of weights adjustment during error minimization
2.	SVM	Max_iteration	It provides maximum iterations for solver for optimization of weights
		C	It is a regularization or penalty parameter to resolve ill-posed optimization complications and improve generalizability of training SVM model
		Gamma	Coefficient of RBF kernel
3.	RF	Kernel	It is a special type of mathematical functions that helps to handle nonlinearity and high dimensionality of input data
		Estimators	In RF, decision trees are utilized as base classifiers or estimators
4.	GB	Max_depth	Maximum depth of decision trees
		Min_samples_split	Minimum number of data samples required at each tree node of RF to split it
		Min_samples_leaf	Minimum data samples required at leaf node of RF
5.	Voting	Estimators	In GB, decision trees are utilized as base classifiers or estimators
		Max_depth	Maximum depth of decision trees
		Min_samples_split	Minimum number of data samples required at each tree node of RF to split it
		Min_samples_leaf	Minimum data samples required at leaf node of RF
6.	Stacking	Base classifiers	Diverse supervised classifiers are selected as estimators and combined in single voting ensemble architecture
		Weights	Weights assigned to different classifiers or estimators
		Voting (hard/soft)	In hard voting setting, voting classifier utilizes majority voting rule for combining the decisions of different base classifiers, whereas soft voting setting uses weighted average probabilities technique for combination
7.	Voting	Combining strategies	It defines the type of combining strategies for voting classifiers
		Base classifiers	Diverse supervised classifiers are selected as estimators and combined in single stacking ensemble architecture
8.	Stacking	Meta-classifier	It is a supervised classifier which processes and combines the outcomes of base classifier as a second computational layer in stacking ensemble architecture

Kindly refer user guide of Scikit-learn Python toolbox for detailed explanation of these parameters

were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Table 8.

References

Adler FJ, Caplan WM, Carlson MP, Goebel ED, Henslee HT, Hicks IC, Larson TG, McCracken MH, Parker MC, Rascoe B, Schramm MW, Wells JS (1971) Future petroleum provinces of the midcontinent. In: Cram IH (ed) Future petroleum provinces of

the United States—their geology and potential. American Association of Petroleum Geologists, Memoir, Tulsa, pp 985–1120

Al-Anazi A, Gates ID (2010) A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Eng Geol* 114(3–4):267–277. <https://doi.org/10.1016/j.enggeo.2010.05.005>

Al-Mudhafar MJ (2017) Integrating well log interpretations for lithofacies classification and permeability through advanced machine learning algorithms. *J Pet Explor Prod Technol* 7:1023–1033. <https://doi.org/10.1007/s13202-017-0360-0>

Anifowose F, Labadin J, Abdulraheem A (2015) Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines. *Appl Soft Comput* 26:483–496. <https://doi.org/10.1016/j.asoc.2014.10.017>

Anifowose FA, Labadin J, Abdulraheem A (2017) Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization. *J Pet Sci Eng* 151:480–487. <https://doi.org/10.1016/j.petrol.2017.01.024>

Aplin AC, Macquaker JHS (2011) Mudstone diversity: origin and implications for source, seal, and reservoir properties in

- petroleum systems. *AAPG Bull* 95(12):2031–2059. <https://doi.org/10.1306/03281110162>
- Avanzini A, Balossino P, Brignoli M, Spelta E, Tarchiani C (2016) Lithologic and geomechanical facies classification for sweet spot identification in gas shale reservoir. *Interpretation* 4(3):SL21–SL31. <https://doi.org/10.1190/int-2015-0199.1>
- Baba K, Bahi L, Ouadif L (2014) Enhancing geophysical signals through the use of Savitzky–Golay filtering method. *Geophys Int* 53(4):399–409. [https://doi.org/10.1016/S0016-7169\(14\)70074-1](https://doi.org/10.1016/S0016-7169(14)70074-1)
- Bestagini P, Lipari V, Tubaro S (2017) A machine learning approach to facies classification using well logs. *SEG Tech Prog Exp Abs*. <https://doi.org/10.1190/segam2017-17729805.1>
- Bhattacharya B, Timothy RC, Pal M (2016) Comparison of Supervised and unsupervised approaches for mudstone lithofacies classification: case studies from the Bakken and Mahantango-Marcellus Shale, USA. *J Nat Gas Sci Eng* 33:1119–1133. <https://doi.org/10.1016/j.jngse.2016.04.055>
- Bhattacharya S, Kavousi P, Carr T, Pantaleone S (2019) Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: a case study from the Marcellus Shale, North America. *J Pet Sci Eng* 176:702–715. <https://doi.org/10.1016/j.petrol.2019.01.013>
- Chaki S, Routray A, Mohanty WK, Jenamani M (2015) A novel multi-class SVM based framework to classify lithology from well logs: a real-world application. In: Annual IEEE India conference (INDICON). <https://doi.org/10.1109/indicon.2015.7443653>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1023/A:1022627411411>
- Diaz MB, Kim KY, Kang TH, Shin HS (2018) Drilling data from an enhanced geothermal project and its pre-processing ROP forecasting improvement. *Geothermics* 72:348–357. <https://doi.org/10.1016/j.geothermics.2017.12.007>
- Dietterich TG (2000) An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach Learn* 40(2):139–157. <https://doi.org/10.1023/A:1007607513941>
- Farshad M, Sadeh J (2014) Transmission line fault location using hybrid wavelet-Prony method and relief algorithm. *Int J Electr Power Energy Syst* 61:127–136. <https://doi.org/10.1016/j.ijepes.2014.03.045>
- Friedman J (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232
- Gifford CM, Agad A (2010) Collaborative multi-agent rock facies classification from wireline well log data. *Eng Appl Artif Intell* 23(7):1158–1172. <https://doi.org/10.1016/j.engappai.2010.02.004>
- Gu Y, Bao Z, Song X, Patil S, Ling K (2019) Complex lithology prediction using probabilistic neural network improved by continuous restricted Boltzmann machine and particle swarm optimization. *J Pet Sci Eng* 179:966–978. <https://doi.org/10.1016/j.petrol.2019.05.032>
- He L, Croix ADL, Wang J, Ding W, Underschultz JR (2019) Using neural networks and the Markov Chain approach for facies analysis and prediction from well logs in the Precipice Sandstone and Evergreen Formation, Surat Basin, Australia. *Mar Pet Geol* 101:410–427. <https://doi.org/10.1016/j.marpetgeo.2018.12.022>
- Healey SP et al (2018) Mapping forest change using stacked generalization: an ensemble approach. *Remote Sens Environ* 204:717–728. <https://doi.org/10.1016/j.rse.2017.09.029>
- Ho TK (1995) Random decision forest. In: Proceedings of the 3rd international conference on document analysis and recognition, Montreal, pp 278–282
- Horrocks T, Holden EJ, Wedge D (2015) Evaluation of automated lithology classification architectures using highly-sampled wireline logs for coal exploration. *Comput Geosci* 83:209–218. <https://doi.org/10.1016/j.cageo.2015.07.013>
- Imamverdiyev Y, Sukhostat L (2019) Lithological facies classification using deep convolutional neural network. *J Pet Sci Eng* 174:216–228. <https://doi.org/10.1016/j.petrol.2018.11.023>
- Jewett JM, Merriam DF (1959) Geologic framework of Kansas—a review for geophysicists. In: Hambleton WW (ed) Proceedings of symposium on geophysics in Kansas, vol 137. Kansas Geological Survey Bulletin, pp 9–52
- Jia J, Yang N, Zhang C, Yue A, Yang J, Zhu D (2013) Object-oriented feature selection of high spatial resolution images using an improved Relief algorithm. *Math Comput Model* 58(3–4):619–626. <https://doi.org/10.1016/j.mcm.2011.10.045>
- Kansas Geological Survey (KGS) (2009). <http://www.kgs.ku.edu/Magellan/Logs/>
- Kittler J, Hafez M, Duin RPW, Matas J (1998) On combining classifiers. *IEEE Trans Pattern Anal Mach Intell* 20(3):226–239. <https://doi.org/10.1109/34.667881>
- Kocberber S, Collins RE (1990) Impact of reservoir heterogeneity on initial distributions of hydrocarbons. In: Proceeding of SPE 65th annual technical conference and exhibition, New Orleans, SPE 20547. <https://doi.org/10.2118/20547-MS>
- Kohavi R (1995) A study of cross validation and bootstrap for accuracy estimation and model selection. *Proc Int Joint Conf Artif Intell* 2:1137–1143
- Kuncheva LI (2004) Combining pattern classifiers: methods and algorithms. Wiley, Hoboken (ISBN: 0-471-21078-1)
- Li Z, Schieber J (2017) Detailed facies analysis of the Upper Cretaceous Tununk Shale Member, Henry Mountains Region, Utah: implications for mudstone depositional models in epicontinental seas. *Sediment Geol* 364:141–159. <https://doi.org/10.1016/j.sedgeo.2017.12.015>
- Liu H, Chen C (2019) Multi-objective data-ensemble wind speed forecasting model with stacked spares autoencoder and adaptive decomposition-based error correction. *Appl Energy* 254:113686. <https://doi.org/10.1016/j.apenergy.2019.113686>
- Ma YZ (2011) Lithofacies clustering using principal component analysis and neural network: applications to wireline logs. *Geoscience* 43:401–419. <https://doi.org/10.1007/s11004-011-9335-8>
- Masoudi P, Tokhmechi B, Bashari A, Jafari MA (2012) Identifying productive zones of the Sarvak formation by integrating outputs of different classification methods. *J Geophys Eng* 9(3):282–290. <https://doi.org/10.1088/1742-2132/9/3/282>
- Merriam DF (1963) The geologic history of Kansas. Kansas Geological Survey, Bulletin 162, 317 p. <http://www.kgs.ku.edu/Publications/Bulletins/162/index.html>
- Moradi M, Tokhmechi B, Masoudi P (2019) Inversion of well logs into rock types, lithofacies and environmental facies, using pattern recognition, a case study of carbonate Sarvak Formation. *Carbonates Evaporites* 34:335–347. <https://doi.org/10.1007/s13146-017-0388-8>
- Mustaffa Z, Yusof Y (2010) A comparison of normalization techniques in predicting dengue outbreak. In: Proceeding of international conference on business and economics research, IACSIT Press, Kuala Lumpur, Malaysia
- Newell KD (1987a) Overview of petroleum geology and production in Kansas. *Kansas Geol Surv Bull* 237. <http://www.kgs.ku.edu/Publications/Bulletins/237/Newell2/overview.pdf>
- Newell KD (1987b) Sub-Chattanooga subcrop map of Salina basin, Kansas. Kansas Geological Survey, Open-file Report 87-4. http://www.kgs.ku.edu/Publications/OFR/1987/OFR87_4/index.html
- Qi L, Carr TR (2006) Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas. *Comput Geosci* 32(7):947–964. <https://doi.org/10.1016/j.cageo.2005.10.020>
- Raeesi M, Moradzadeh A, Ardejani FD, Rahimi M (2012) Classification and identification of hydrocarbon reservoir lithofacies and

- their heterogeneity using seismic attributes, logs data, and artificial neural networks. *J Pet Sci Eng* 82–83:151–165. <https://doi.org/10.1016/j.petrol.2012.01.012>
- Ribeiro GT, Mariani VC, Coelho LDS (2019) Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Eng Appl Artif Intell* 82:272–281. <https://doi.org/10.1016/j.engappai.2019.03.012>
- Santos ROV, Vellasco MMBR, Artola FAV, da Fontoura SAB (2003) Neural net ensembles for lithology recognition. In: Windeatt T, Roli F (eds) *Multiple classifier systems, MCS 2003. Lecture notes in computer science*, vol 2709. Springer, Berlin. https://doi.org/10.1007/3-540-44938-8_25
- Sebtosheikh MA, Salehi A (2015) Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir. *J Pet Sci Eng* 134:143–149. <https://doi.org/10.1016/j.petrol.2015.08.001>
- Sesmero MP, Ledezma AI, Sanchis A (2015) Generating ensembles of heterogeneous classifiers using stacked generalization. *Wiley Interdiscip Rev Data Min Knowl Discov* 5(1):21–34. <https://doi.org/10.1002/widm.1143>
- Shalabi LA, Shaaban Z, Kasasbeh B (2006) Data mining: a preprocessing engine. *J Comput Sci* 2(9):735–739. <https://doi.org/10.3844/jcssp.2006.735.739>
- Skurichina M, Duin RPW (2001) Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal Appl* 5(2):121–135. <https://doi.org/10.1007/s100440200011>
- Spain DR, Merletti GD, Dawson W (2015) Beyond volumetrics: unconventional petrophysics for efficient resource appraisal (example from the Khazzan field, Sultanate of Oman). In: *Proceedings of 5th SPE Middle East unconventional resources conference and exhibition, Muscat, Oman*. <https://doi.org/10.2118/172920-MS>
- Suarez-Alvarez MM, Pham DT, Prostov MY, Prostov YI (2012) Statistical approach to normalization of feature vectors and clustering of mixed datasets. *Proc R Soc A Math Phys Eng Sci* 468(2145):2630–2651. <https://doi.org/10.1098/rspa.2011.0704>
- Tewari S, Dwivedi UD (2018a) A novel automatic detection and diagnosis module for quantitative lithofacies modeling. In: *Proceeding of Abu Dhabi international petroleum conference and exhibition, UAE, SPE-192747*. <https://doi.org/10.2118/192747-MS>
- Tewari S, Dwivedi UD (2018b) Ensemble-based big data analytics of lithofacies for automatic development of petroleum reservoirs. *Comput Ind Eng* 128:937–947. <https://doi.org/10.1016/j.cie.2018.08.018>
- Tewari S, Dwivedi UD (2019) A real world investigation of TwinSVM classifier for classification of petroleum drilling data. In: *Proceeding of IEEE TENSYPMP 2019 symposium, Kolkata, India*
- Tewari S, Dwivedi UD, Shiblee M (2019) Assessment of big data analytics based ensemble estimator module for the real time prediction of reservoir recovery factor. In: *Proceeding of SPE Middle East oil and gas show and conference held in Manama, Bahrain, SPE-194996-MS*, pp 18–21. <https://doi.org/10.2118/194996-MS>
- Urbanowicz RJ, Meeker M, Cava ML, Olson RS, Jason HM (2017) Relief-based feature selection: introduction and review. *J Biomed Inform* 85:189–203. <https://doi.org/10.1016/j.jbi.2018.07.014>
- Wang G, Carr TR (2012) Methodology of organic-rich shale lithofacies identification and prediction: a case study from Marcellus shale in the Appalachian basin. *Comput Geosci* 49:151–163. <https://doi.org/10.1016/j.cageo.2012.07.011>
- Wang Y, Wang D, Geng N, Wang Y, Yin Y, Jin Y (2019) Stacking based ensemble learning of decision trees for interpretable prostate cancer detection. *Appl Soft Comput* 77:188–204. <https://doi.org/10.1016/j.asoc.2019.01.015>
- Windeatt T (2006) Accuracy/diversity and ensemble MLP classifier design. *IEEE Trans Neural Netw* 17(5):1194–1211. <https://doi.org/10.1109/TNN.2006.875979>
- Wolpert DH (1992) Stacked generalization. *Neural Netw* 5(2):241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Xie Y, Zhu C, Zhou W, Li Z, Liu X, Tu T (2018) Evaluation of machine learning methods for formation lithology identification: a comparison of tuning process and model performance. *J Pet Sci Eng* 60:182–193. <https://doi.org/10.1016/j.petrol.2017.10.028>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.