Petroleum Science 20 (2023) 540–558

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Gas liquid cylindrical cyclone flow regime identification using machine learning combined with experimental mechanism explanation

Zhao-Ming Yang [a], Yu-Xuan He [a], Qi Xiang [a], Enrico Zio [b, c], Li-Min He [d], Xiao-Ming Luo [d], Huai Su [a, *], Ji Wang [a], Jin-Jun Zhang [a, **]

[a] *National Engineering Laboratory for Pipeline Safety/MOE Key Laboratory of Petroleum Engineering/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, 102249, Beijing, China*
[b] *MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France*
[c] *Dipartimento di Energia, Politecnico di Milano, Via La Masa 34, 20156, Milano, Italy*
[d] *College of Pipeline and Civil Engineering, China University of Petroleum-East China, 266580, Qingdao, China*

ARTICLE INFO

Article history:
Received 10 May 2022
Received in revised form
13 September 2022
Accepted 14 September 2022
Available online 4 October 2022

Edited by Xiu-Qiu Peng

Keywords:
Gas liquid cylindrical cyclone
Machine learning
Flow regimes identification
Mechanism explanation
Algorithms

ABSTRACT

The flow regimes of GLCC with horizon inlet and a vertical pipe are investigated in experiments, and the velocities and pressure drops data labeled by the corresponding flow regimes are collected. Combined with the flow regimes data of other GLCC positions from other literatures in existence, the gas and liquid superficial velocities and pressure drops are used as the input of the machine learning algorithms respectively which are applied to identify the flow regimes. The choosing of input data types takes the availability of data for practical industry fields into consideration, and the twelve machine learning algorithms are chosen from the classical and popular algorithms in the area of classification, including the typical ensemble models, SVM, KNN, Bayesian Model and MLP. The results of flow regimes identification show that gas and liquid superficial velocities are the ideal type of input data for the flow regimes identification by machine learning. Most of the ensemble models can identify the flow regimes of GLCC by gas and liquid velocities with the accuracy of 0.99 and more. For the pressure drops as the input of each algorithm, it is not the suitable as gas and liquid velocities, and only XGBoost and Bagging Tree can identify the GLCC flow regimes accurately. The success and confusion of each algorithm are analyzed and explained based on the experimental phenomena of flow regimes evolution processes, the flow regimes map, and the principles of algorithms. The applicability and feasibility of each algorithm according to different types of data for GLCC flow regimes identification are proposed.

1. Introduction

The gas liquid cylindrical cyclone (GLCC) is popularly and widely used in gas-liquid two-phase separation, especially in natural gas fields and transportation station. Since gas-liquid separation is a necessary part of natural gas development and transportation, the flow regime identification of the GLCC has been more important for the flow assurance. With the development of machine learning, the clustering, classification and prediction based on the data are easier and more accurate to be realized. Since the characteristics of flow regime can be represented by different kinds of data, machine learning, which is a technology driven by data, has been combined with multiphase flow regime identification.

1.1. Gas liquid cylindrical cyclone (GLCC)

Gas liquid cylindrical cyclone is a kind of multiphase separator with high efficiency and compactness (Shoham and Kouba, 1998; Kouba et al., 2006; Kulkarni and Shinde, 2016). The main structure of GLCC consists of a vertical cylinder and tangential inlets. As the mixture of gas and liquid enters the GLCC through the tangential inlet, then they form a cyclone and are separated by the difference

* Corresponding author.
** Corresponding author.
E-mail addresses: suhuai@cup.edu.cn (H. Su), zhangjj@cup.edu.cn (J.-J. Zhang).

https://doi.org/10.1016/j.petsci.2022.09.005

of centrifugal force. The liquid phase and gas phase will exist through the upper and downer outlets.

The researches of GLCC mainly focus on the structure and efficiency optimization (Kouba et al., 1995; Arpandi et al., 1996; Chirinos et al., 2000; Movafaghian et al., 2000; Rosa et al., 2001; Yue et al., 2019; Moncayo et al., 2018), including how to combine with other separator to be an integrated system (Peixoto et al., 2005; Ju et al., 2010; Iyer et al., 2010; Mogseth, 2008; Khoi Vu et al., 2009). The flow regimes can reflect the continuity of the phase flowing in pipes or wells, which is important to flow assurance. However, as the safety getting much more important, the flow regime in GLCC start to be focused.

1.2. Flow regime in GLCC

The research of flow regimes of GLCC is not as advanced as the research of flow regimes in pipes, as the flow fields of GLCC include horizon pipe flow, vertical pipe flow, incline pipe flow and cyclone flow. The flow regime identification includes direct method and indirect method. The former is mainly based on the observation of experimental phenomena or the flow regimes pictures, which is also the main method of researching GLCC flow regimes. Based on different section of GLCC, the flow regimes are different in definitions and patterns. For the inlet of GLCC, four flow regimes are defined as smooth flow, stratified flow, slug flow, annular flow in the inlet, and three flow regimes included churn flow, annular flow and ribbons flow in liquid carry-over (LCO) (Hreiz et al., 2014). For the flow regimes in LCO, based on different GLCC structures, there are also classification of swirling film flow, churn flow, and ribbons flow (Xu et al., 2018), or swirling annular flow, interim flow and complete churn flow (Wang et al., 2021). In practical fields, the inlet of GLCC is connected to a vertical pipe, based on the effects of vertical pipe, the inlet flow regimes are classified by continuous flow, intermittent flow and annular flow (Yang et al., 2022). The classifications of flow regimes of GLCC are concluded in Table 1.

Flow regimes in GLCC can reflect the flow condition, however, in practical industry, the flow regimes cannot be observed like those in laboratory, so the indirect method of identification is more reliable. Therefore, it is necessary and important to identify the flow regime indirectly for flow assurance. The flow regimes of GLCC are identified by the liquid film thickness (Wang et al., 2021) and pressure drops of GLCC (Yang et al., 2022). Yang et al. have applied statistic methods to identify the flow regimes and separation efficiencies of GLCC using pressure drops, and the corresponding relationships between flow regimes and pressure drops have also been analyzed. The data used for flow regime identification are in forms of flow regime photos, sensors signal, optical probe signals et al. There is a problem to consider that some data used for flow regimes identification are difficult or impossible to be measured in practical fields, because the measurement equipment or sensor in practical fields are not in the same level as those in laboratory, such as high-speed camera, particle image velocimetry, impendence probes and radiation equipment. As a result, when choosing the data as the input of machine learning, it is necessary to use the data which are easy to acquire in practical fields, such as pressure drops or velocities.

There is still no function which can classify flow regimes clearly or describe the boundary of flow regimes clearly, neither in the multiphase flow of pipeline nor GLCC. Since the flow regimes identification is important but difficult, the machine learning technology which is driven by data has been more popular in the flow regimes identification.

1.3. Machine learning and its application in flow regime

As a core part of Artificial Intelligence (AI), machine learning has been developed and proceeded rapidly. The essence of machine learning is to extract the characteristics of a group of data, then to realize the function of clustering, classifying or predicting (Jordan and Mitchell, 2015; Pedregosa et al., 2011; Lecun et al., 2015). The machine learning can be classified with supervised learning and unsupervised learning, the former type of which needs to be labeled with the properties of each data group and the latter dose not. Data for machine learning are mainly divided into training group and testing group. For the flow regimes, the data which can represent different flow regimes has been used in kinds of algorisms. The algorithms of machine learning applied in flow regimes researches are shown in the mind map of Fig. 1 and the types of data are shown in mind map of Fig. 2. It is hard to say that there is the best algorithm to identify the flow regimes, because the accuracy of the algorithm is based on the data amount, the types of data and the quality of data. The data of flow regimes can be classified with parameter data, signal data and image data. In addition, most of the data for the algorithm input is from the experiments, however, the experimental instruments are not equipped in all practical environments, especially for some with high price and strict requirement of environment. As described in section 1.2, it is necessary and important to consider the availability of the practical fields data which is used for machine learning. It also means the machine learning algorithms should have higher accuracy based on these kinds of data which is easier to acquire from practical fields. The algorithms for the flow regimes are following the mainstream of machine learning such as CNN, SVM et al., according to the kinds of input data. From the researches of Fig. 1, few research compared different algorithms of machine learning for the flow regimes identification and explained the mechanism of machine learning algorithms or the reasons for the suitableness of the algorithms. From the types of data of Fig. 2, it can be concluded that most researchers ignored the availability of the data from practical fields. For the research of machine learning algorithms, the variety of the types of input data is meaningful. However, for the application for the engineering, the input data which cannot be acquired from the practical fields do not make much sense.

Except for the advantages of machine learning, the most applications are still for the recommendations for customers, and pushing the information. However, the applications of machine learning for the industry are still in the initial phase, although some technologies in machine learning are well developed. The main

Table 1
Flow regimes in GLCC.

Researcher	Year	Position	Classification
Hreiz et al.	2014	Inclined inlet	smooth flow stratified flow slug flow annular flow
		LCO	churn flow annular flow ribbons flow
		Lower part	bubbly vortex flow excavated vortex flow deeply excavated vortex flow
Xu et al.	2018	LCO	swirling film flow churn flow ribbons flow
Wang et al.	2021	LCO	swirling annular flow interim flow complete churn flow
Yang et al.	2022	Inlet with vertical pipe	continuous flow intermittent flow annular flow

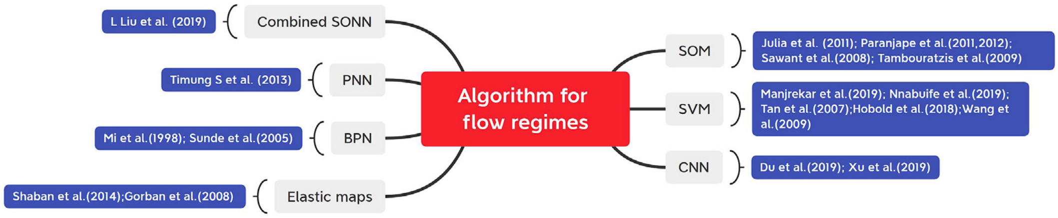

Fig. 1. Algorithms for flow regimes identification. Liu and Bai, 2019a, 2019b, Timung and Mandal, 2013, Mi et al., 1998, Sunde, 2005, Shaban and Tavoularis, 2014, Gorban et al., 2008, Julia et al., 2011, Paranjape et al., 2011a, 2011b, 2011c, 2012, Sawant et al., 2008, Tambouratzis and Pá zsit, 2009, Manjrekar and Dudukovic, 2019, Nnabuife et al., 2019, Tan et al., 2007, Hobold and da Silva, 2018, Wang and Zhang, 2009, Du et al., 2018, Xu et al., 2019.

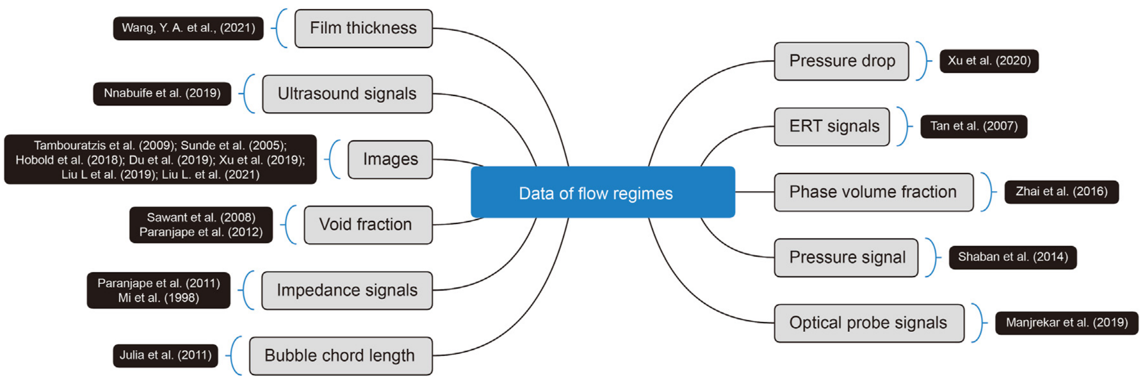

Fig. 2. Types of flow regimes data. Wang et al., 2021, Nnabuife et al., 2019, Tambouratzis and Pá zsit, 2009, Sunde, 2005, Hobold and da Silva, 2018, Du et al., 2018, Xu et al., 2019, Liu and Bai, 2019a, 2019b, Liu et al., 2021, Sawant et al., 2008, Paranjape et al., 2012, 2011a, 2011b, 2011c, Mi et al., 1998, Julia et al., 2011, Xu et al., 2020, Tan et al., 2007, Zhai et al., 2016, Shaban and Tavoularis, 2014, Manjrekar and Dudukovic, 2019.

reason for that is the lack of explanation of the results of machine learning, so it is important to explain the results of the machine learning, experimentally or physically. For the flow regimes in GLCC, few research applies machine learning to identify the flow regimes because the definition of flow regimes in different positions of GLCC have not been unified and the data of flow regimes are not enough. The gaps are concluded of the research of gas-liquid flow regime identification using machine learning:

- The main highlights of most related researches are focused on innovation of experimental methods such as the improved probe method or improved imaging technique, so as to be combined with machine learning algorithms. But some of them focus on the explanation of the results, especially the identification confusion.
- The innovation of types of data and experimental methods ignore the applicability for practical fields and the difficulty of acquiring data from practical fields.
- The machine learning applied for flow regime identification is mainly based on multiphase flow in vertical pipes, circular pipes, and other chemical engineering equipment such as reactors, research of GLCC flow regime identification is still in initial phase, still less the application of machine learning.

Based on the gaps of the current researches, the differences between this paper and other works can be concluded as 1) comparing multiple machine learning algorithms for flow regimes identification after unifying the flow regimes of GLCC in different positions; 2) using the data for machine learning which can be acquired feasibly from practical fields; 3) explain the confusion of flow regimes identification by experimental phenomena and mechanism of algorithms.

This paper integrates the flow regimes in different position of GLCC, including the inlet part with a vertical pipe from our experiments, LCO part, inclined inlet part, and lower part from other literatures, and uses the data of these flow regimes as the input of machine learning. The classical machine learning algorithms are applied for the GLCC flow regimes identification.

The objectives of this study are:

- to explore the feasibility and applicability at the first phase of machine learning algorithms in GLCC flow regimes identification, and choosing the suitable algorithms
- to research the GLCC flow regimes identification using machine learning based on the data which can be easily acquired in practical fields
- to research the reasons for the success and confusion of each algorithm in identification of flow regimes, combined with the algorithm mechanism and experimental explanation.

and the novelties of this paper are:

- to apply machine learning for GLCC flow regimes based on the velocities and pressure drops which can be easily acquired from practical field
- explain the success and confusion of the identification results in view of experimental phenomena, the relationship between algorithms mechanism and data types.

2. Experimental methods

2.1. Experimental setup

The experimental setup is based on the GLCC testing system of (Yang et al., 2022). Since in the practical field, there is a vertical pipe in front of the inlet of GLCC, a vertical pipe is set in the experiment.

The whole experimental setup and key sizes are shown in Fig. 3. With the diameter of 220 mm and height of 2345 mm, the GLCC prototype is similar to the real one in practical fields.

Four modules are set in the system including measurement module, power module, image module and data acquisition module. The measurement module is for gas volume flow, liquid mass flow and pressure measurement. Based on the parameters of fluids and pipes, the superficial velocities of gas and liquid are calculated. The power module is used to provide power to gas and liquid by screw compressor and centrifugal pump. Flow regimes are recorded by camera in image module. After the data are measured by measurement module, the data acquisition module transfers and records the signals and data. In Fig. 3, the gas flow is shown in green and the liquid is shown in bule. All the details of the equipment and process are shown in Table 2 and (Yang et al., 2022), and the characteristics range for different flow patterns and fluids type are shown in Table 3. In the experiment, flow regimes are researched in five different inlet structures. The difference of each inlet structure is the necking ratio r_n, which is the ratio of the gas-liquid flow area to the inlet pipe section area.

2.2. Data reduction

Two kinds of data which can reflect flow regimes are acquired in experiments. It is necessary to clarify that the gas and liquid flow rate (superficial velocity) are converted by the measurement of quantity of gas/liquid flow. The identification of GLCC flow regimes is based on practical fields project, and before the mixture of gas and liquid entering GLCC, the quantity of gas/liquid/mixture flow is measured by multiphase flow meter, both in overland and underwater gathering & transportation engineering. For some gas fields in China, the quantities of gas and liquid flow are possible to be measured like measuring pressure drops, with error range in ±3%. The pressure drop is the time-varying kind of data which is periodical in each flow regime, as shown in Eq. (1). As the differences of flow regimes are based on the different slug presences and moving characteristics, and the slug presences and moving characteristics result in the different characteristics of pressure drops, so the pressure drops can reflect the flow regimes. The other kind of data is superficial velocities of gas phase and liquid phase, as shown in Eqs. (2) and (3). The data for flow regimes identification are calculated by the data which are measured by measurement module. The flow regimes in different velocities and inlets are shown in Fig. 4.

$$\Delta p = \mathrm{P}_1 - \mathrm{P}_2 \tag{1}$$

$$v_g = \frac{Q_g}{A} = \frac{4Q_g}{\pi d^2} \tag{2}$$

$$v_l = \frac{Q_l}{A} = \frac{4Q_l}{\pi d^2} \tag{3}$$

For GLCC in practical fields, the vertical pipe is set for researching the effects of the vertical pipe, so the pressure measuring point P1 is set in front of the vertical pipe. In order to remove the probable effects of sampling time, a dimensionless time is used as

$$t = \frac{T_S - T_0}{t_0} \tag{4}$$

where T_s is the testing time, T_0 is the initial time of one test, and t_0 is a testing period.

The data used in flow regimes identification are pressure drops and superficial velocities. The flow regimes identification using the two types of data is based on the classical machine learning. For the flow regimes in lack of velocity data or the ranges of their velocities are quite far away from others, these kinds of flow regimes are not chosen as the researched objectives. The algorithms of machine learning are shown in Fig. 7. In ensemble model, the basic is Decision Tree model, which is based on a tree structure. The bifurcation and pruning of the tree are determined by some calculation index, and the classification of the object is determined. The multi-classification of the predicted value is realized by multiple bifurcation refinement classification. The Decision Tree model consists Bagging Tree and Boosting Tree. The Bagging Tree and Random

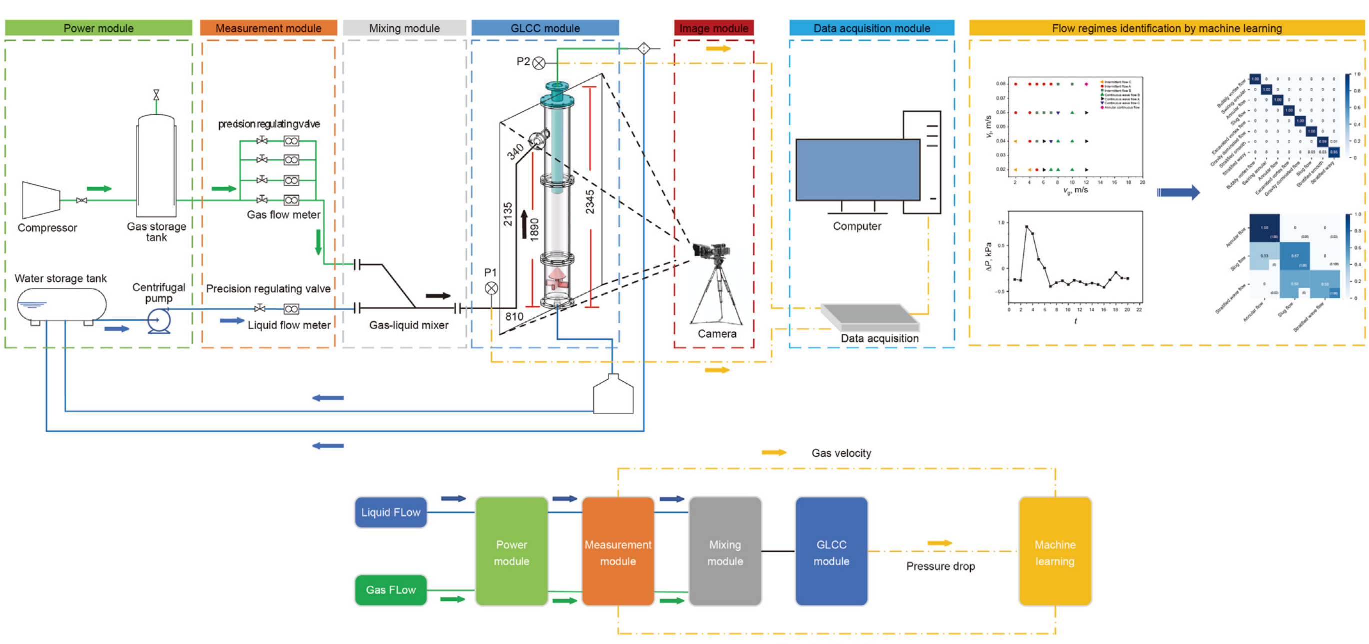

Fig. 3. Experimental setup.

Table 2
Experiment module and device.

Module Name		Device Name	Range	Precision	Producer
Measurement Module	Gas Measurement	orifice plate flowmeter(EJA115 0.864 mm)	0.11–0.77, Nm^3/h	±5%	Yokogawa Sichuan Instrument Co.
		orifice plate flowmeter(EJA115 2.527 mm)	0.876–6.3, Nm^3/h	±5%	Yokogawa Sichuan Instrument Co.
		orifice plate flowmeter(EJA115 6.350 mm)	5.34–37.8, Nm^3/h	±5%	Yokogawa Sichuan Instrument Co.
		vortex shedding flowmeter	30-400, Nm^3/h	±1%	Yokogawa Shanghai Instrument Co.
	Liquid Measurement	Mass flowmeter(DY015-DN15)	0–27200, kg/h	±0.75%	Micro Motion Co.
	Pressure Measurement	Rosemount3595	adjustable	±5%	Rosemount Co.
		Keller PA23	0–21000, bar	±0.3%	KELLER AG fur Druckmesstechnik
		Rosemount3051	0–62, kPa	±5%	Rosemount Co.
Power Module		screw compressor	–	–	Atlas Co.
		IH single stage centrifugal pump	–	–	Yangzijiang pump Co.
Image Module		Camera X–S10	–	–	FUJIFILM (China) Investment Co.
Data Acquisition Module		NI PCI-6229&Labview	–	–	National Instrument Co.

Table 3
Characteristics range for different flow patterns and fluids type.

Flow pattern	v_g range, m/s	v_l range, m/s	*GLR*
Annular flow	>12	>0.06	125–150
Continuous flow	8–12	0.02–0.06	133.3–600
Intermittent flow	2–8	0.02–0.06	25–300
Fluid type	Density, kg/m^3	Viscosity, Pa·s	Temperature, °C
water	997.048	1.005×10^{-3}	20
air	1.205	1.790×10^{-5}	20

Forest are similar type of machine learning. The key of Bagging Tree and Random Forest is introducing randomization into its construction procedure and then making an ensemble of it, which aims to increase the variety of tree types and training set.

The other type of tree model is Boosting Tree, including Adaboost, GBDT, XGBoost and LGBM. The main characteristic of Adaboost is that it can change the weights of incorrectly classified terms in order to focus more about these cases. GBDT is used for regression originally and then for classification after optimization, however, this algorithm can be more complex with the increase of data. XGBoost can be seen as the improvement of GBDT by considering the complexity of tree and it has improved the efficiency of calculation. The proposal of LGBM is to solve the problem when GBDT meets with large amounts of data, and it has successfully make the GBDT faster than before.

Besides the algorithms based on Decision Tree, the other classical algorithms for classification are also applied. SVM uses the support vector to determine the boundary of classification. By the limitation of its mechanism, SVM needs to use one-to-others order to realize the multi-classification. KNN is an algorithm which also has the function of classification and regression, but is different from the ensemble models, because it just records and remember all the data. The advantage of KNN is that it can describe complex boundary between the classifications. The algorithms above realize the classification by voting or statistics from different sub models, but Bayesian Model uses probability to classify. The Bernoulli Bayes classifier and Gaussian Bayes classifier from Bayesian Model are used for the flow regimes identification. These two sub models assume the input is as the distribution of Bernoulli and Gaussian. The MLP model is also called ANN, and uses multi layers to classify the objective data, and it is the most complex algorithm of all, since the layers of itself and nodes are still unknown.

3. Results and discussion

3.1. GLCC inlet flow regimes with a vertical pipe

The GLCC inlet flow regimes with a vertical pipe are found in the type of annular flow, intermittent flow including three patterns, and continuous flow including three patterns, which are classified in A, B and C. Comparing the different classifications mean different evolution processes. Combining the flow regimes phenomena with the corresponding pressure drops, the changes in pressure drops are related to the evolution processes, especially to the slug formation, moving, releasing and other motions. The flow regimes evolution process and pressure drops are shown in Fig. 5, and described in detail by (Yang et al., 2022). The presence of a vertical pipe can significantly form a slug inside and have the effects on the flow regimes development. It has been found that the pressure drops change periodically, because in fixed gas and liquid superficial velocities, the process of corresponding flow regime is periodical.

For the flow regimes in this experiment, the corresponding v_g and v_l range are shown in Table 3, and the ratio of v_g to v_l is also shown.

3.2. Flow regimes of GLCC

The flow regimes in different positions of GLCC are shown in Fig. 6. In the LCO, there are 3 flow regimes which are swirling annular flow, interim flow and complete churn flow (Wang et al., 2021). In the lower part of GLCC, the flow regimes are bubbly vortex flow, excavated vortex flow and deeply excavated vortex flow (Hreiz et al., 2014). When the inlet is inclined, the flow regimes are defined as stratified smooth flow, slug flow, stratified wavy flow and annular flow (Hreiz et al., 2014), however, the research has not considered the vertical pipe in front of that. The experiment in this paper has added the horizon inlet with the vertical pipe, and the flow regimes are continuous flow, intermittent flow and annular flow.

3.3. Flow regimes identification

The flow regimes are often classified by gas and liquid superficial velocity, in the form of flow regimes map, so velocities can be used as the standard of flow regimes identification. All the flow regimes in the four positions have their related flow regime maps, so the velocity data samples can be obtained by the maps. Meanwhile, in the experiment of this paper, it has been found that pressure drops may be the characteristics to identify flow regimes, so the pressure drops are also used as the input data of machine learning, also by the same algorithms of Ensemble models, SVM, KNN, Baysian Model and MLP, as shown in Fig. 7 (Rebentrost et al., 2014; Massaoudi et al., 2021; Wang et al., 2021; Abellán and Masegosa, 2012).

The 12 machine learning algorithms include linear classification (KNN), nonlinear classification (MLP), high-dimensional spatial

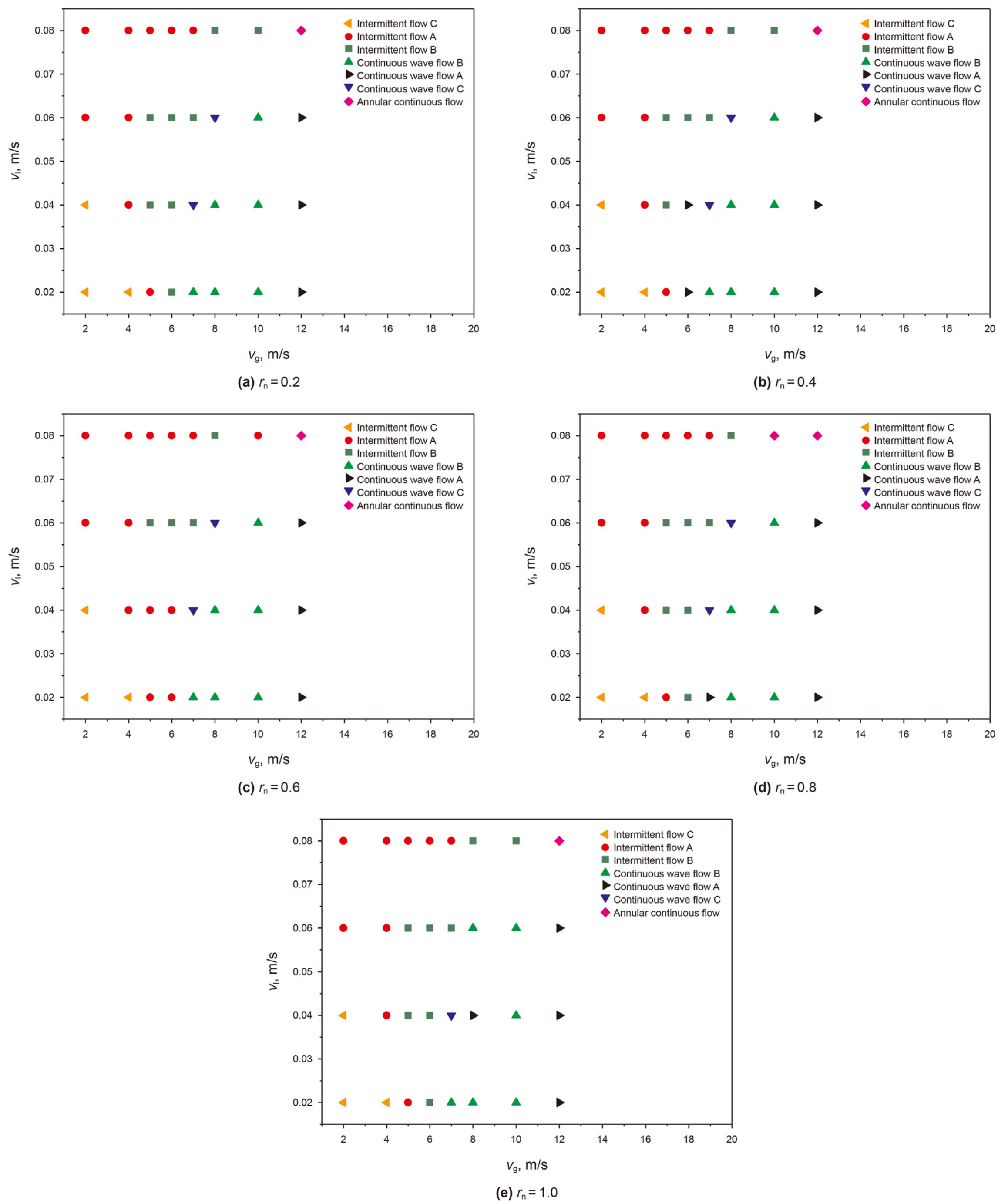

Fig. 4. Flow regimes in different inlets and velocities.

mapping (SVM), Tree-based classification and its advancements (Decision Tree, Bagging Tree, RandomForest, Adaboost, XGBoost, LGBM), Bayesian decision model (Bernoulli Bayesian Network and Gaussian Bayesian Network). As Fig. 7 shows, gas/liquid superficial velocities and pressure drops are chosen as the objective data type to realize GLCC flow regimes identification. The gas/liquid superficial velocities can be seen as two-dimensional data without changing with time, and the pressure drop is a kind of temporal data, which can be easily acquired from practical fields. According to the types of data, the mainstream kinds of machine learning for identification, including Tree Model & Ensemble Model, SVM, KNN, Bayesian Model and MLP. The five models above cover the current machine learning models with different principles as comprehensively as possible. It can give support for quick selection of models in practical application.

Both pressure drop data and superficial velocity data are classified into training set and test set. There are 5091 groups of data (Wang et al., 2021; Hreiz et al., 2014) which contains gas superficial velocity and liquid superficial velocity, and 1026 groups of data of pressure drops. Each group of data has been marked with its corresponding flow regimes. The ratio of training set to testing set is 8:2 and the choosing of data for these two sets is random in each identification.

The most popular indexes to measure the results of machine

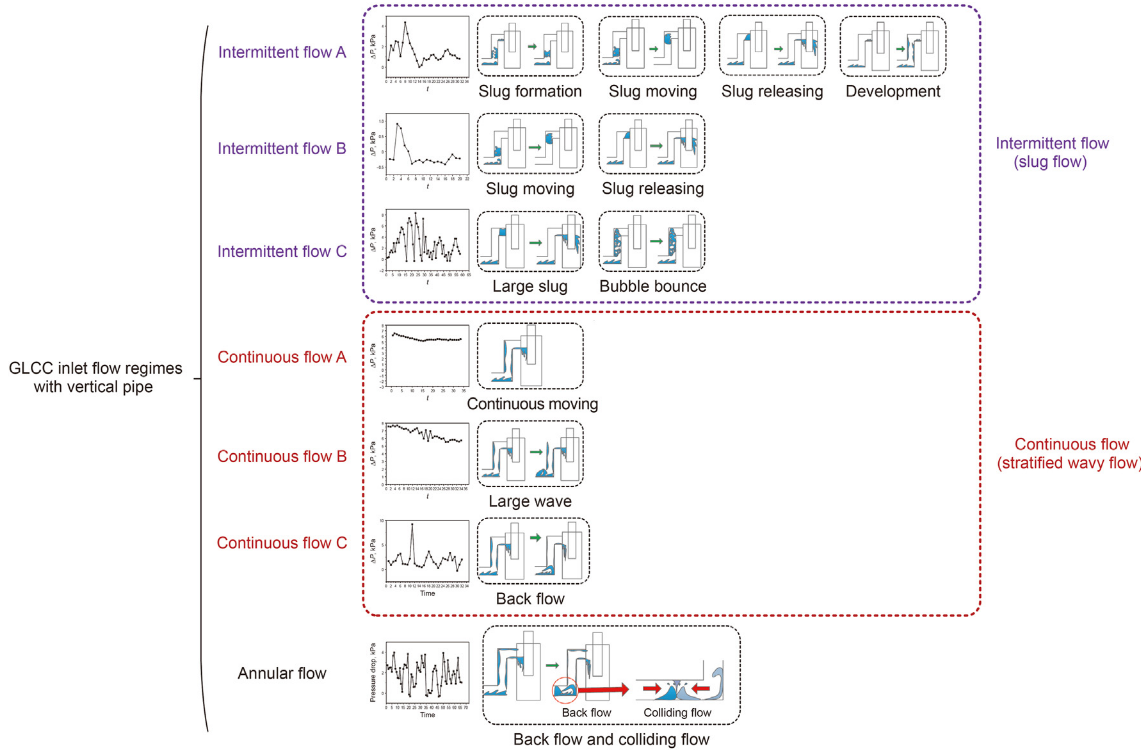

Fig. 5. GLCC flow regimes with vertical pipe.

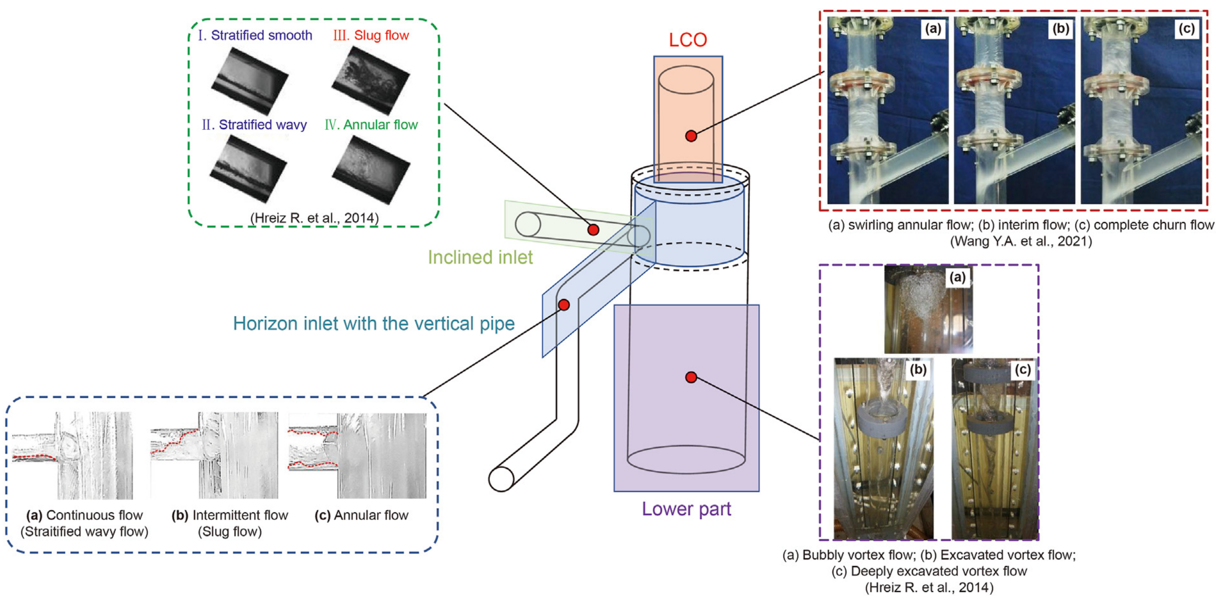

Fig. 6. Flow regimes of GLCC.

learning identification are accuracy, confusion matrix and Receiver Operating Characteristic (ROC). The definition of accuracy is

$$accuracy = \frac{N_{\mathrm{correct}}}{N_{\mathrm{total}}} \tag{5}$$

where N_{correct} is the number of data which are identified correctly, and N_{total} is the number of the total corresponding testing data. The accuracy is the most simple measurement index. The confusion matrix is a matrix combining real results and output results. Each column of the confusion matrix represents the prediction category, and the total number of each column represents the number of data predicted for this category. Each row represents the true category of data. The ROC curve is another index to measure the classification results. The abscissa of ROC is False Positive Rate(FPR), and the ordinate is True Positive Rate(TPR). The FPR and TPR are shown as:

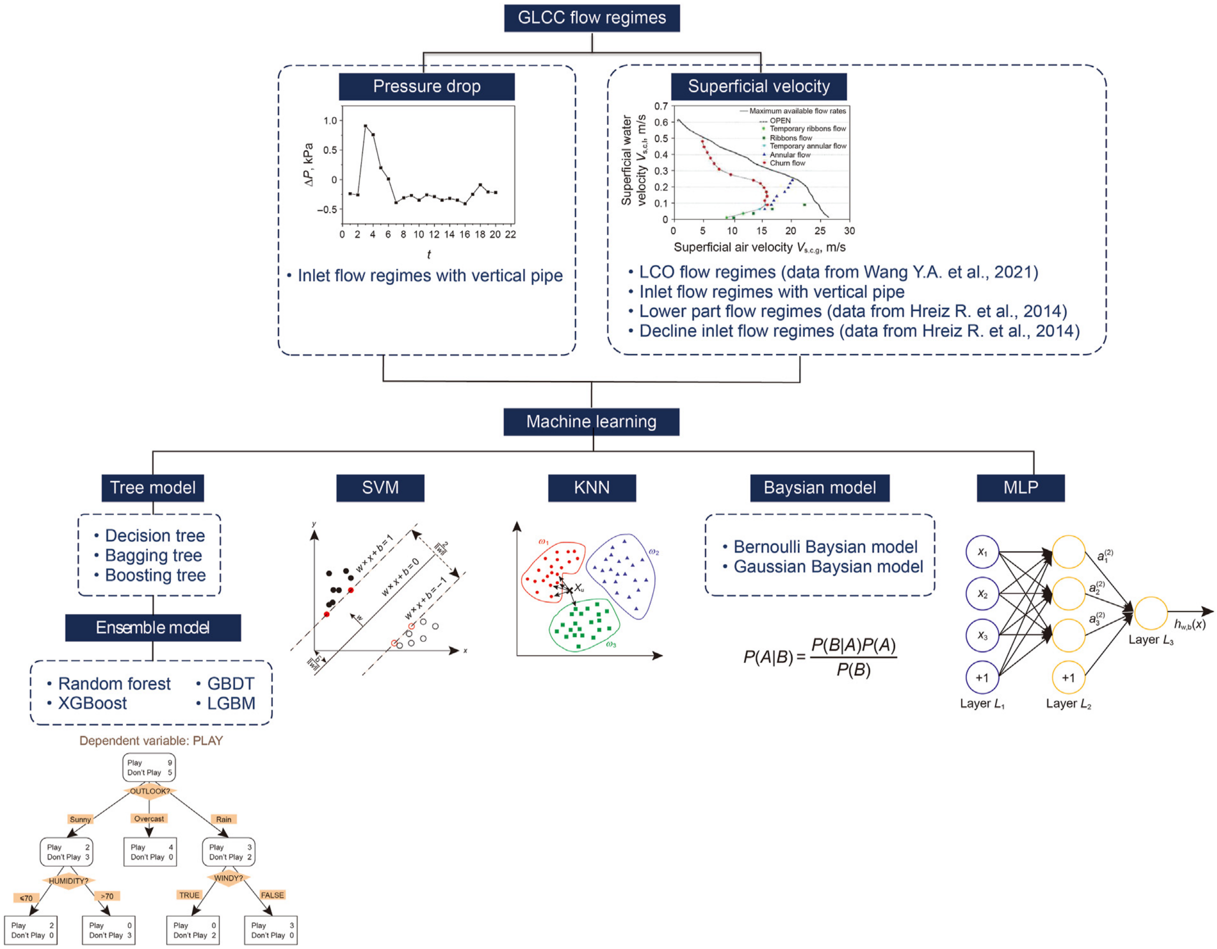

Fig. 7. Flow regimes and machine learning.

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

where the classifications are assumed to be positive(P) and negative(N), and the results are true(T) and false(F), so TP is true positive, FP is false positive, TN is true negative and FN is false negative. The area of ROC curve can reflect the accuracy of identification, and when the area is near 1.0, the more accurate the results of identification are.

The 12 algorithms can be initially selected by Table 4. ○ means the corresponding algorithm has higher accuracy for superficial velocity in the initial election, and △ means higher accuracy for pressure drops. The applicability of each algorithm will be analyzed in the following parts. The accuracies of 12 algorithms are shown in Table 4, and it can be seen that for flow regimes identification by superficial velocities, the Random Forest, XGBoost, GBDT, Bagging Tree, Decision Tree, LGBM and MLP can identify the flow regimes with the accuracy above 0.99, so the algorithms based on decision tree are suitable for the flow regimes identification by velocity data. For the flow regimes identification by the pressure drop, the

Table 4
Accuracy of algorithms for flow regime identification.

Algorithm	Accuracy by superficial velocity	Accuracy by pressure drop	Primary election
Random forest	0.9971	0.8182	○△
KNN	0.9253	0.4545	○
SVM	0.5501	0.3636	–
Adaboost	0.7868	0.6363	–
XGBoost	0.9951	1	○△
GBDT	0.9961	0.8182	○△
Bagging Tree	0.9941	1	○△
Decision Tree	0.9961	0.6363	○
Bernoulli Baysian Network	0.4126	0.6363	–
Gaussian Baysian Network	0.8988	0.6363	○
LGBM	0.9941	0.8182	○△
MLP	0.9941	0.8182	○△

XGBoost and Bagging Tree can identify all the flow regimes by the pressure drops.

3.3.1. Identification by superficial velocity

Fig. 8 shows the confusion matrix of regimes identification with superficial velocities. The results of confusion matrix fit the results of accuracy. It can be seen from confusion matrix that for the algorithms with accuracy above 0.99, there is confusion in stratified wavy flow, slug flow and stratified smooth flow. Few stratified wavy flows are identified with stratified smooth flow and slug flow, and the reasons for that are: 1) there are few points in the transition area between stratified wavy flow-stratified smooth flow, and stratified wavy flow-slug flow. To avoid the effects of transition area, the data sampling is uniformly-spaced. However, when the sampling data point is located in transition area, like Fig. 9, the classification boundary of flow regimes described by machine

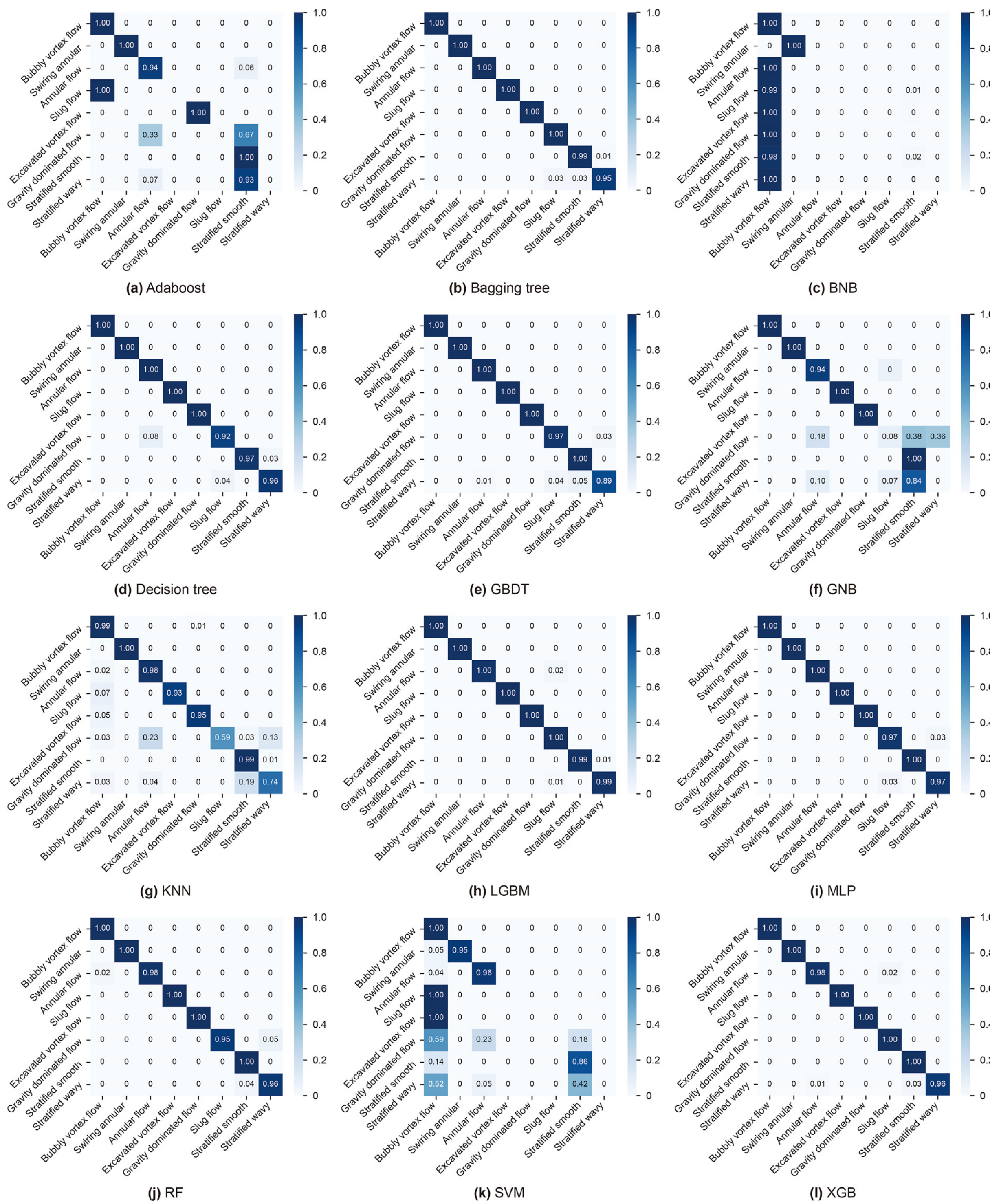

Fig. 8. Confusion matrix of regime identification with superficial velocities.

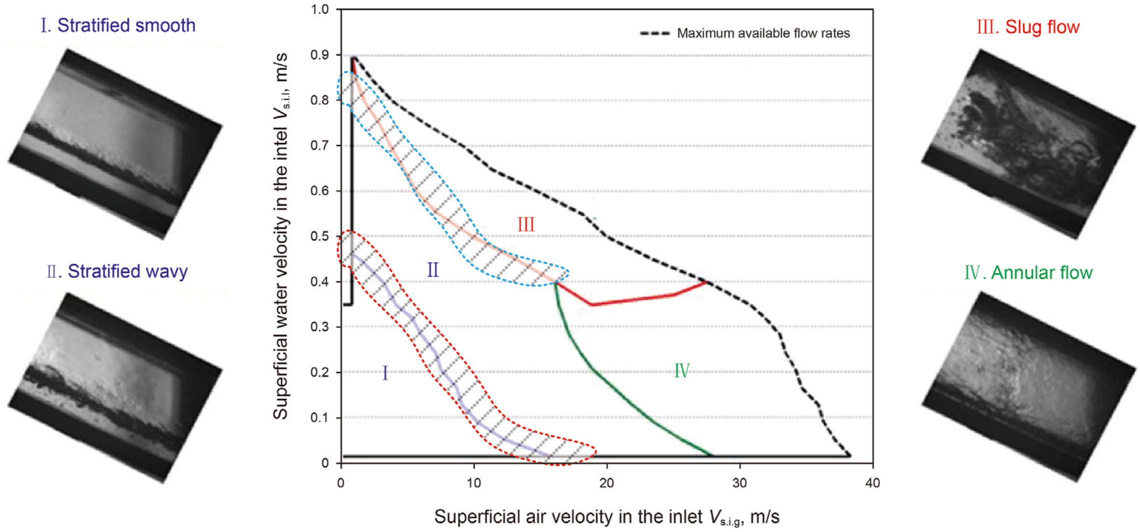

Fig. 9. Inclined inlet flow regimes map of GLCC.

learning algorithms may be not the same as the real boundary in flow regimes map, which can affect the identification results of machine learning. Actually, in experiment, the boundary of flow regimes cannot be described accurately and clearly, for which the real boundary is an area, so the sampling points has not been selected more than the other areas. 2)For the unification of inlet flow regimes of horizon inlet and incline inlet, the flow regimes in the two inlets are mixed and trained by each machine learning algorithm, so when the superficial velocities are the same, the identification may be confused by few data.

Based on the Fig. 13, the ROC curves are in accord with the results of confusion matrix. For the algorithms with accuracy above 0.99, the ROC curve of the flow regimes can reach the area larger than 0.99, which mean the results of identification are accurate. It can be concluded from all the machine learning algorithms that the ensemble models and MLP algorithms are accurate for the GLCC flow regimes identification.

In detail, the ensemble models are based on the decision tree algorithms, including Bagging Tree, Decision Tree, GBDT, LGBM. The details of the ensemble models are shown in Fig. 10 and Fig. 11. According to the brief introduction of ensemble models of machine learning, the process of flow regimes identification by boosting algorithm is shown in Fig. 10. Firstly, the data are sampled from the total database randomly, each of which has been labeled by *position*, *flow regime*, v_{sl}, v_{sg}. When the sampling time is *T*, the amount of training sets is *T*, and we can get *T* base learning algorithms. Based on the ensemble strategies, such as weights method, all the base learning algorithms are ensembled to form a more powerful algorithm. The key of Boosting algorithm is if the flow regime is identified wrong, as shown in red, the weights of the wrong item will be added. The aim of that is to make the final powerful algorithm can learn the difficult samples better.

For the Bagging algorithm, the difference of that from boosting algorithm is that the base learning algorithms are less related and are trained at the same time, as shown in Fig. 11. The ensemble method of these algorithms is also different, such as average method and voting method. In addition, for the application for classification or identification, the ensemble method is chosen as voting method. The accuracy of bagging method depends on the stableness of base learning algorithms. The most famous boosting algorithms are AdaBoost and GBDT.

The other important algorithm is decision tree, whose learning process can be shown in Fig. 12. The decision tree can be seen as a tree structure. Each internal node represents a judgment on an attribute, each branch represents the output of a judgment result, and finally each leaf node represents a classification result. The key of decision tree for flow regime identification is the thresholds which are set to classify the flow regimes. For the identification by superficial velocities, the thresholds mean that the algorithm should find the critical gas and liquid superficial velocities to distinguish the flow regimes, just like describing the flow regimes map.

According to the characteristics of the algorithms described above, the ensemble models and decision tree can be combined with decision tree algorithm, so as to acquire other ensemble algorithms, such as

$$\text{Bagging} + \text{Decision Tree} = \text{Random Forest} \tag{8}$$

$$\text{Adaboost} + \text{Decision Tree} = \text{Boosting Decision Tree} \tag{9}$$

$$\text{Gradient Boosting} + \text{Decision Tree} = \text{Gradient Boosting Decision Tree (GBDT)} \tag{10}$$

As shown in Fig. 8 and Table 4, AdaBoost algorithm shows the terrible accuracy in flow regimes identification by superficial velocities, but Decision Tree can effectively improve the accuracy of flow regimes identification. The fundamental machine learning algorithms such as Decision Tree and Bagging Tree can identify the flow regimes by superficial velocities accurately, so the other ensemble algorithms can. This has also shown that: 1) superficial velocities are suitable indexes to characterize and identify the flow regimes. 2) For the fundamental algorithms like Boosting algorithm and Bagging algorithm, the Bagging algorithm is better than the Boosting algorithm. 3) Machine learning can effectively identify most flow regimes by the velocity data. 4) The fundamental algorithms of machine learning can realize the identification of flow regimes, and it is unnecessary to use the deep learning. 5) For the accuracy of Decision Tree, the key point of flow regimes identification is to determine the threshold.

In addition, there are still some algorithms which are not suitable for the flow regimes identification. For the Adaboost algorithm, it will give too many weights to the samples which are difficult to classify, so this results in that the imbalance of data can

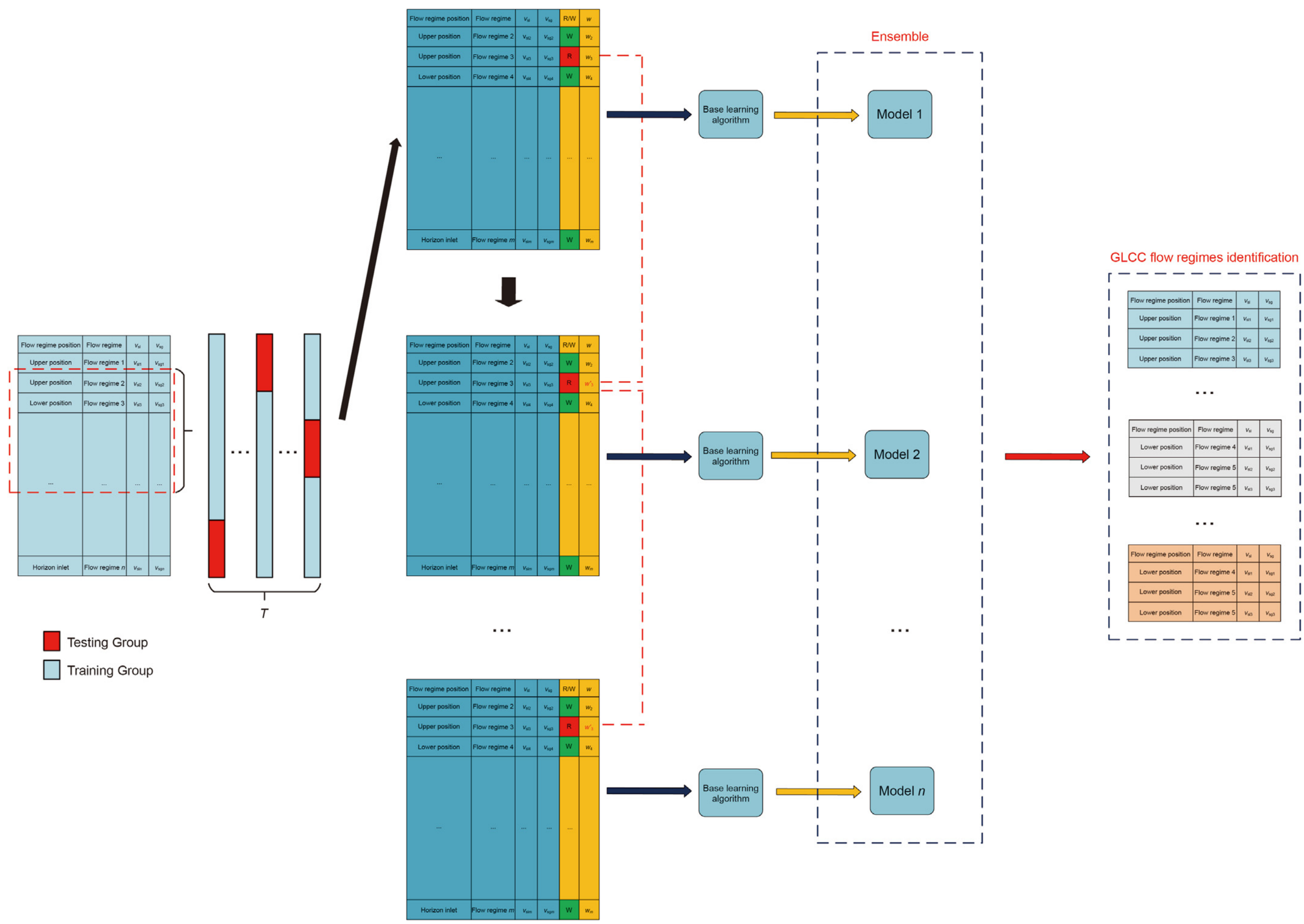

Fig. 10. Boosting algorithm.

decrease the accuracy of identification decrease. The data points of superficial velocities for flow regimes are actually not distributed in the same number for each regime, so the AdaBoost algorithm is not suitable for this identification. That means the numbers of sampling points for each flow regime are different. For example, as Fig. 4d shows, there are 3 flow regimes including 7 evolution processes, and the sampling points of each process are marked by different shapes and colors. It can be seen from Fig. 4d that different flow regimes have different number of sampling points, which are called data points in machine learning algorithms. Since Adaboost is sensitive to the imbalance data groups, the difference of sampling points numbers in each flow regime can affect the accuracy of Adaboost. For the classical SVM algorithm, the best area is dichotomy, but the flow regimes identification is not a simple dichotomy problem, so the classical SVM is not suitable for the flow regimes identification neither. For the Bayesian Network algorithm, based on the essence of this kind of algorithm, which depends on the probabilities to classify, require that the characteristic properties have no relationship, if not, the Bayesian Network algorithm will be ineffective.

3.3.2. Identification by pressure drops

For the pressure drop identification by pressure drops, it has been found the results of identification are different in different tests, and Fig. 14 shows the 12 algorithms to identify the flow regimes 100 times and combine the confusion matrixes of highest and lowest accuracies into one confusion matrix. It can be seen from the new Fig. 14 that XGBoost and GBDT have the accuracy in range of 0.856–1.000 (slug flow), 0.91–1.00 (stratified wave flow) and 0.92–1.00 (slug flow), 0.895–1.00 (stratified wave flow). The reasons for 100% accuracy are that: 1) the results of machine learning are random, including accuracy of 100%, so the results from two tests cannot show the accuracy of each algorithm for the flow regime identification; 2) compared with the data from practical fields, there are much less noise in experimental data. Different from superficial velocities, the pressure drops change with time periodically, and the sample data amount is less than superficial velocities. For this type of data, few ensemble algorithms show the advantages in identification. It can be seen from the identification results after multi-tests, the most confusions are some of the slug flow can be identified as stratified wave flow, and some annular flow can be identified as slug flow. The ROC curves in Fig. 15 show the similar results of the flow regime identification.

Combined with Fig. 14, the Table 5 and Fig. 16 show the averages of accuracies and variances of 100 random tests. GBDT and XGBoost have the identification accuracy above 0.90, however, the accuracy of identifications by pressure drops are not as good as those by superficial velocities, and the identifications by superficial velocities are more stable than those by pressure drops.

Based on the results of the identification from machine learning, the combination of slug flow A and stratified wave flow C, which are the different evolution processes of slug flow and stratified

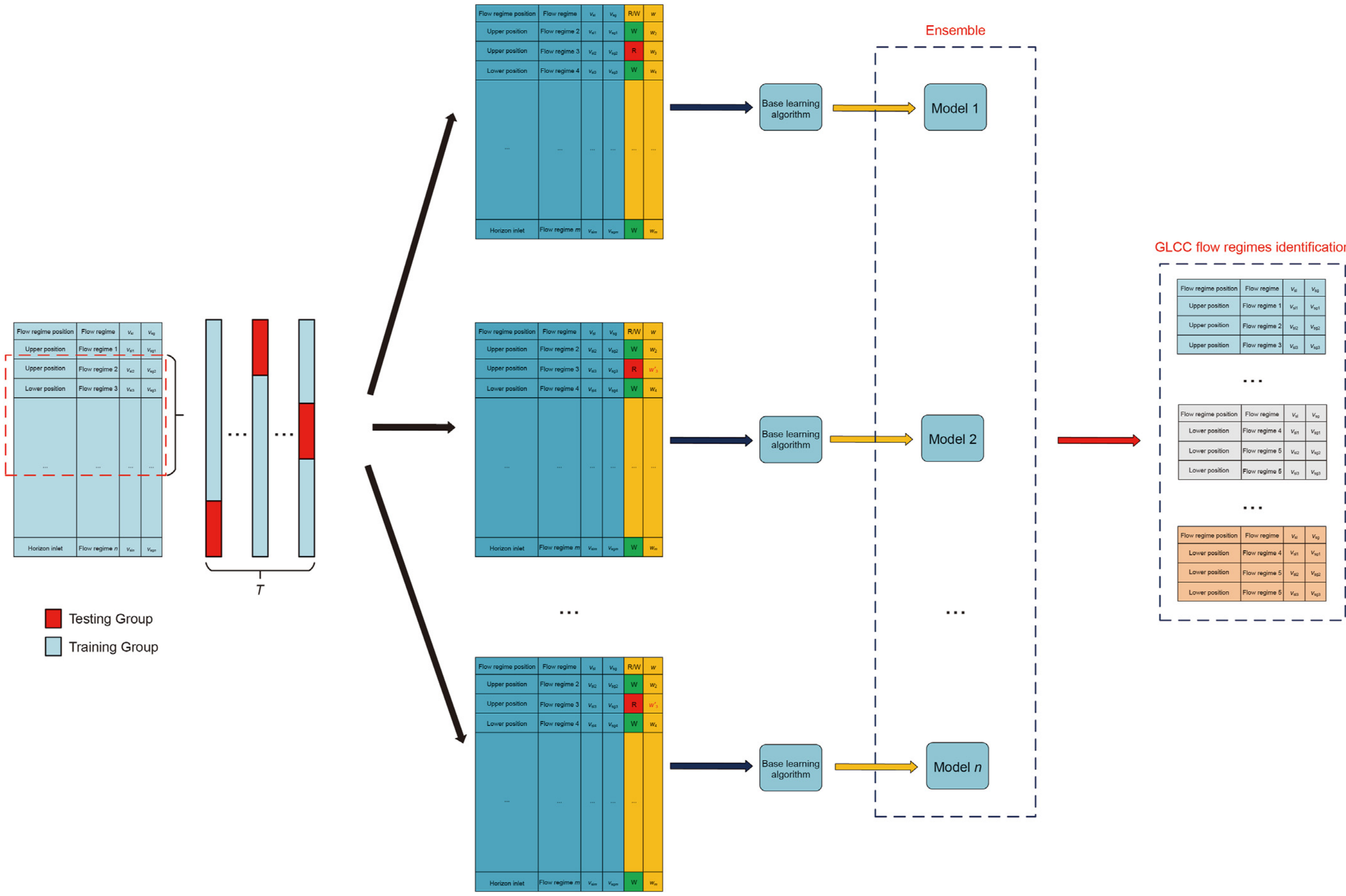

Fig. 11. Bagging tree.

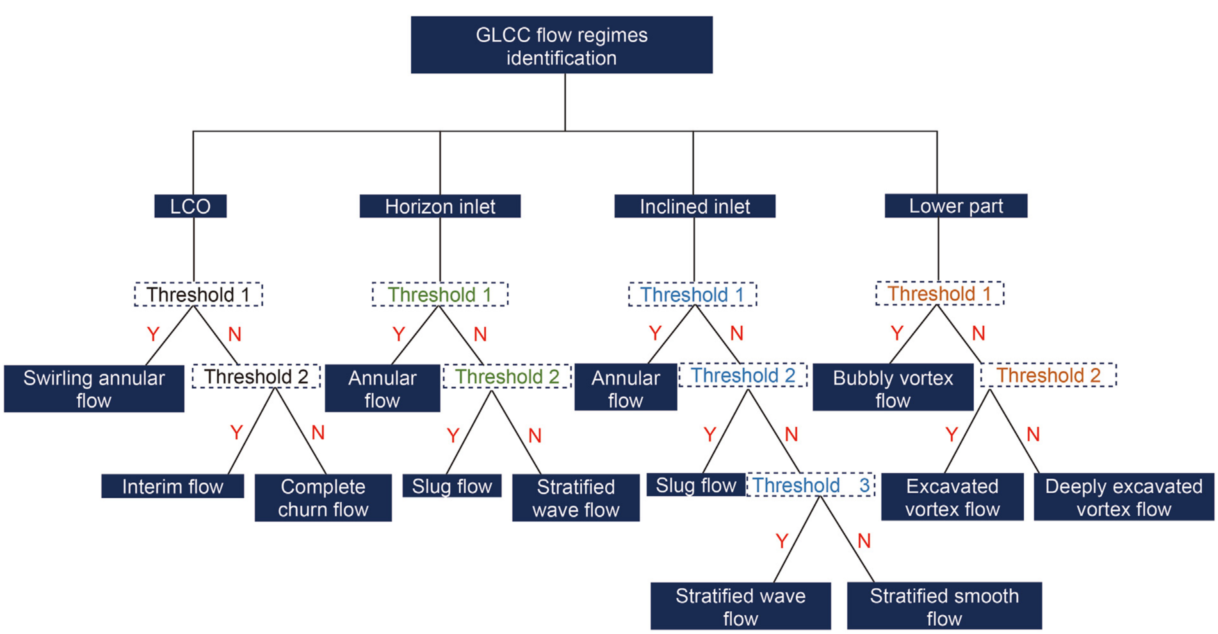

Fig. 12. Decision tree.

wave flow; and the combination of slug flow C and annular flow, which is the evolution of slug flow, may confuse the algorithms of identification. The explanation of the results of machine learning has been recognized as the difficulty, especially for the confusion of the results. That's also the reason for that the process of machine learning is called Black Box. In this paper, experimental results, the

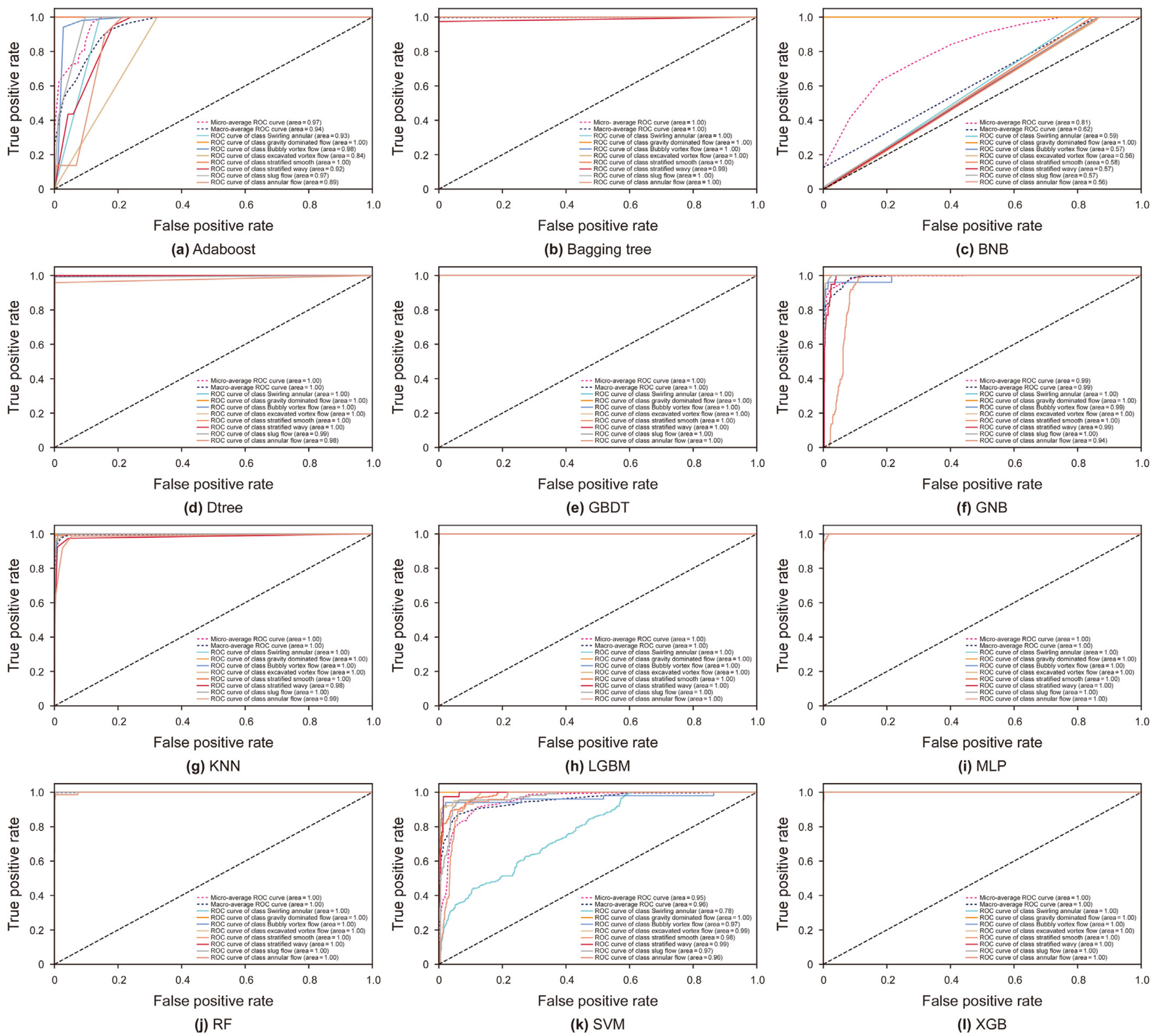

Fig. 13. ROC of regime identification with superficial velocities.

evolution process of flow regimes and the characteristics of algorithms are used to explain the confusion of identification results.

Based on the experiments of horizon inlet flow regimes, there are three types of slug flow and three types of stratified wave flow, which are distinguished by A, B, and C. As can be seen from Fig. 17, the pressure drops of slug flow A and stratified wave flow C have similar trends. Comparing the process of flow regime and the synchronous pressure drops, the increase of pressure drops for slug flow A is the result of slug moving. With the slug moving in the vertical pipe, the pressure drop keeps increasing. Then the slug arrives at the inlet, and release into the GLCC, meanwhile the pressure drops begin to decrease. There is also a period when the pressure drops in stable fluctuations, and this is because the liquid is accumulating to form the slug. In addition, the slug cannot form easily after each accumulating process, however, from the whole perspective of slug flow A, its pressure drops and evolution process are periodical. For stratified wave flow C, it can be seen from Fig. 17 that there is a process of springback when the liquid flows in the horizon pipe before the vertical pipe. It has been found that not all the velocities can result in the back flow, which is special for the stratified wave flow C and annular flow. The backflow can also have the similar effect which block off the horizon pipe to make the pressure drop increase. That is why the pressure drops of these two flow regimes have the similar trend, which confuses the identification algorithms. Due to the higher gas superficial velocity, the slug cannot form in stratified wave flow C, and high gas superficial velocity is also the reason for that the pressure drops of stratified wave flow are higher than that of slug flow.

For slug flow C and annular flow, the two flow regimes also confuse the other algorithms. Both of these two flow regimes have the larger fluctuations in their pressure drops than other flow regimes. As shown in Fig. 18, the liquid slug consists large amount of gas phase, in the form of the slug containing a lot of bubbles. The volume and length of the slug are larger and longer than other flow

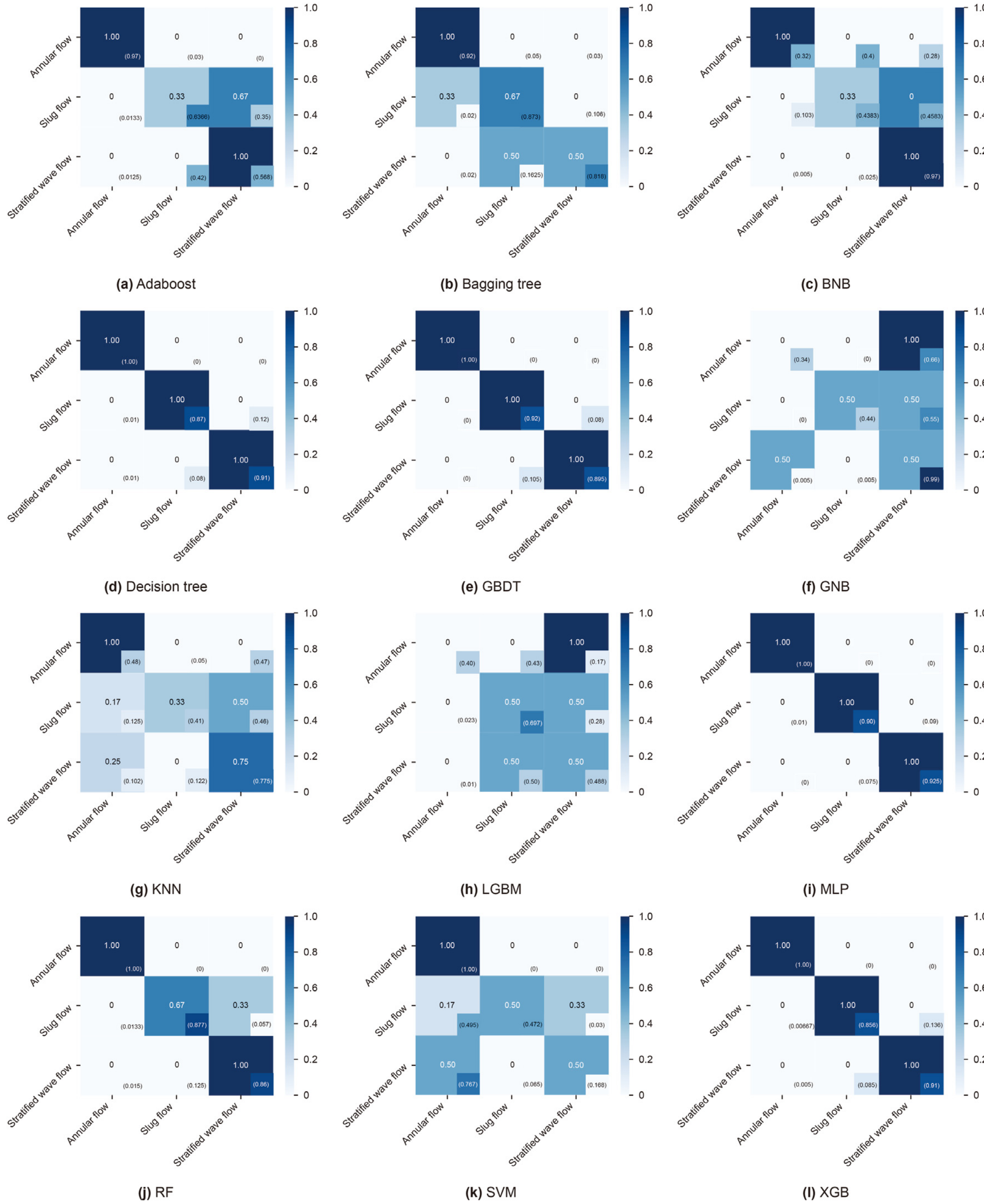

Fig. 14. Confusion matrix of regime identification with superficial velocities.

regimes, and it has been found that for slug flow C, this type of slug can reciprocate in the vertical pipe. The reciprocating motion is also the formation process of liquid slug. If the liquid slug form successfully after the reciprocating motion, a large slug will enter GLCC and release, with the pressure drop increase roughly. There are also some conditions that the large slug does not form after the reciprocating motion, which result in the larger fluctuation of pressure drops but without sudden increase. For the annular flow, there is a springback process in the horizon pipe before the vertical pipe, which is similar with stratified wave flow C. The difference between annular flow and stratified wave flow C is that the springback process for annular flow consists the drastic colliding between two flows in inverse directions. The installation site of the pressure sensor where P_1 is in Fig. 3, is on the upside of the horizon pipe. The colliding of liquid flow can make the liquid reach the pressure sensor, so as to result in that the pressure drops have large fluctuation, which are similar with slug flow C. That can also explain the confusion of slug flow and annular flow.

For the flow regimes identification based on the pressure drop which is changing with time, the GBDT and XGBoost have shown the advantages. The mechanism of Boosting tree has been introduced in 3.3.1. The XGBoost algorithm is actually an improvement

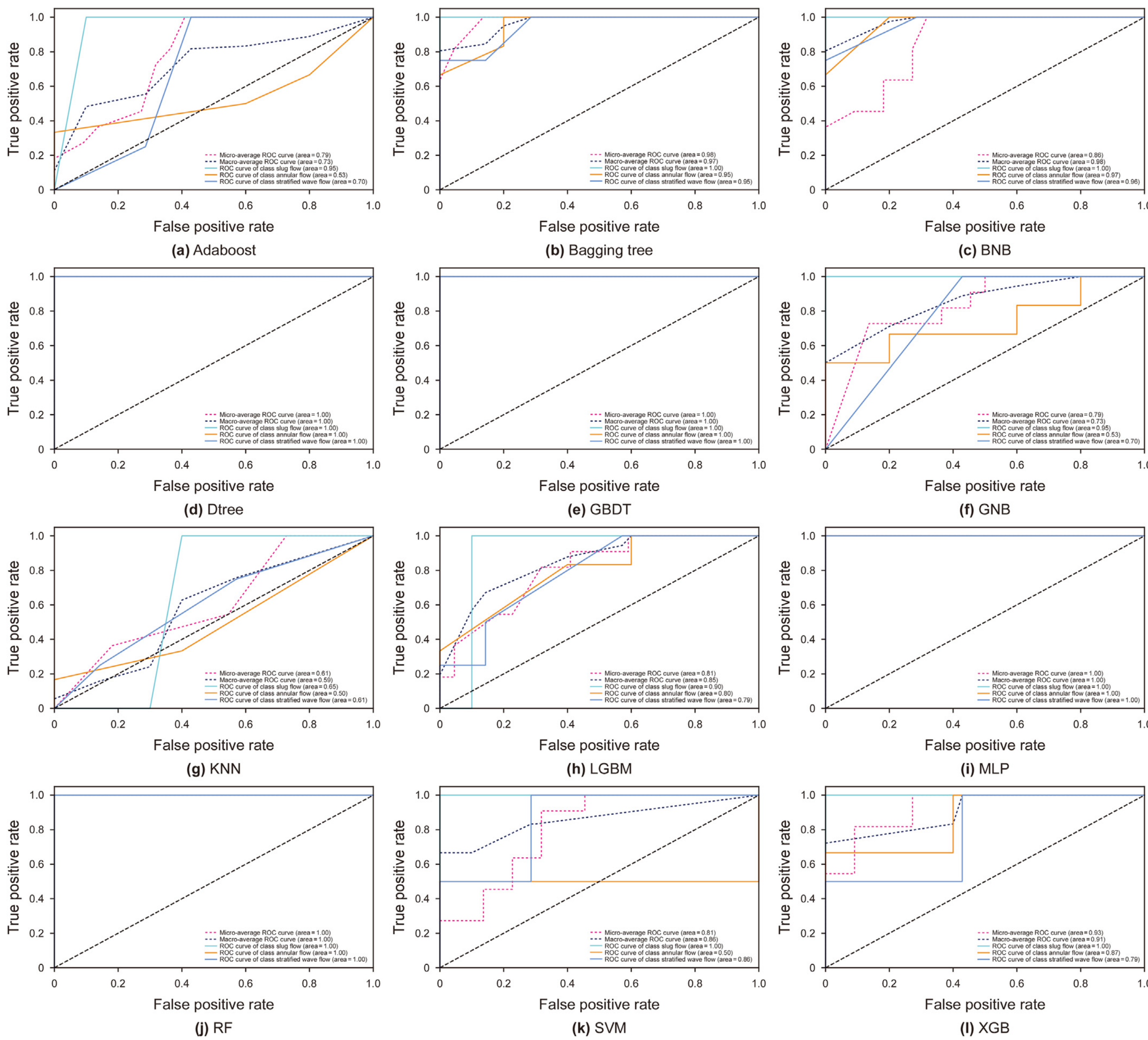

Fig. 15. ROC of regime identification with pressure drops.

Table 5
Average accuracies and variances of GLCC flow regimes identification by pressure drops.

algorithm	accuracy	variance
Adaboost	0.637273	0.008598
GBDT	0.916364	0.010966
BNB	0.605455	0.012392
GNB	0.631818	0.012918
XGB	0.905455	0.013009
BaggingTree	0.882727	0.014909
MLP	0.898182	0.015574
LGBM	0.602727	0.015637
DTree	0.894545	0.016148
SVM	0.386364	0.016425
RF	0.898182	0.01691
KNN	0.563636	0.023875

of GBDT, which is famous with high efficiency of calculation and the easy application in engineering fields. In reality, the base learning algorithm of XGBoost and GBDT is also decision tree, and XGBoost has optimized its sampling method and loss function. Since the pressure drops data are the type of data which is changing with time, such time-varying data are mostly used for prediction in machine learning, but few in identification. From the results of Figs. 14 and 15, it can be concluded the pressure drops are not the best choice of machine learning to identify flow regimes, as most machine learning algorithms have low accuracy in identification. It should be explained here that the 'low' accuracy is a concept compared with the accuracy of identification using superficial velocities. The accuracies of the 12 machine learning algorithms by pressure drops are lower than those by superficial velocities, however, the accuracies of GBDT and XGBoost are still above 0.90 during the repeated tests. The better performance of XGBoost and GBDT are not contradictory to the analysis of the reasons for the

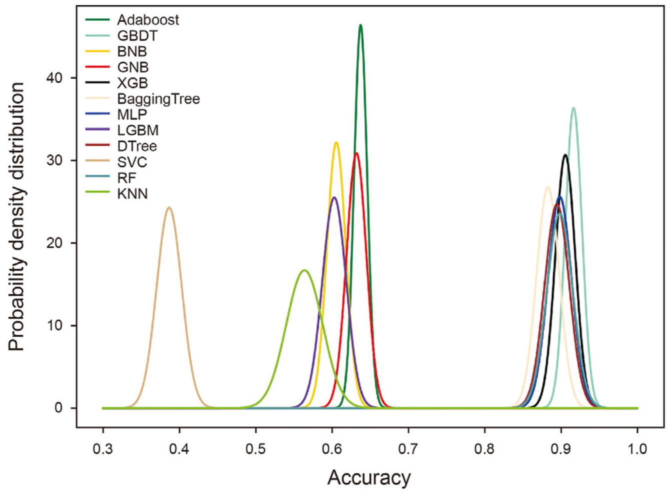

Fig. 16. Probability density distribution of flow regimes identification by pressure drops.

lower accuracy of identification by pressure, which is compared with the accuracy by superficial velocities.

Based on the experimental mechanism explanation above, the complexness of multiphase flow increases the possibility of confusion, especially the vertical pipe is present. In addition, the experiments to measure the pressure drop costs much longer time than those to measure other physical quantities. In result, the data of pressure drops and corresponding flow regimes are much less than the velocity data. Since the research of pressure drops and flow regimes are few, the related literatures which can provide data for reference are few, which is also the reason for the identification results fluctuation in Fig. 14.

Based on the accuracy of GBDT and XGBoost, these two algorithms based on boosting tree, which is based on Decision Tree algorithms are suitable for the flow regimes identification by the pressure drops data, which also have high accuracy during the identification using superficial velocities. In addition, the successful method of a statistic method in GLCC flow regimes identification by (Yang et al., 2022), is the same with the essence of decision tree, which continues to set the thresholds step by step to classify the flow regimes one by one.

It can be concluded that the decision tree algorithm is the core of the GLCC flow regimes identification. However, less machine learning algorithms are suitable for the flow regimes identification by using the pressure drops data. The reasons for that can be concluded that: 1) the data of pressure drops of GLCC are complex. Compared with the superficial velocity, the pressure drops of each flow regime are consisted with dozens of or hundreds of pressure drop points, but the superficial velocity data of each flow regime are only two data points, including gas superficial velocity and liquid superficial velocity. In addition, the complexity of multiphase flow and the presence of vertical pipe have made the pressure drops more complex. 2) The pressure drops of GLCC are changing with the time, which are time-varying and different from other type of data. Although the pressure drops are periodical, they still can confuse the training process of algorithm easily, which are related to the division of the period and the classification boundary of the wave shapes of pressure drops. 3) the amount of pressure drops data of GLCC, which can be labeled by the corresponding flow regime, is few in the experiments of this paper and in other literatures. 4) The types of machine learning algorithms, which can realize the identification based on complex pressure drops data, are limited.

For the algorithms not suitable for the flow regime identification by pressure drops data, the disadvantages of SVM, Adaboost and Baysian Network are introduced in 3.3.1, and they are not accurate

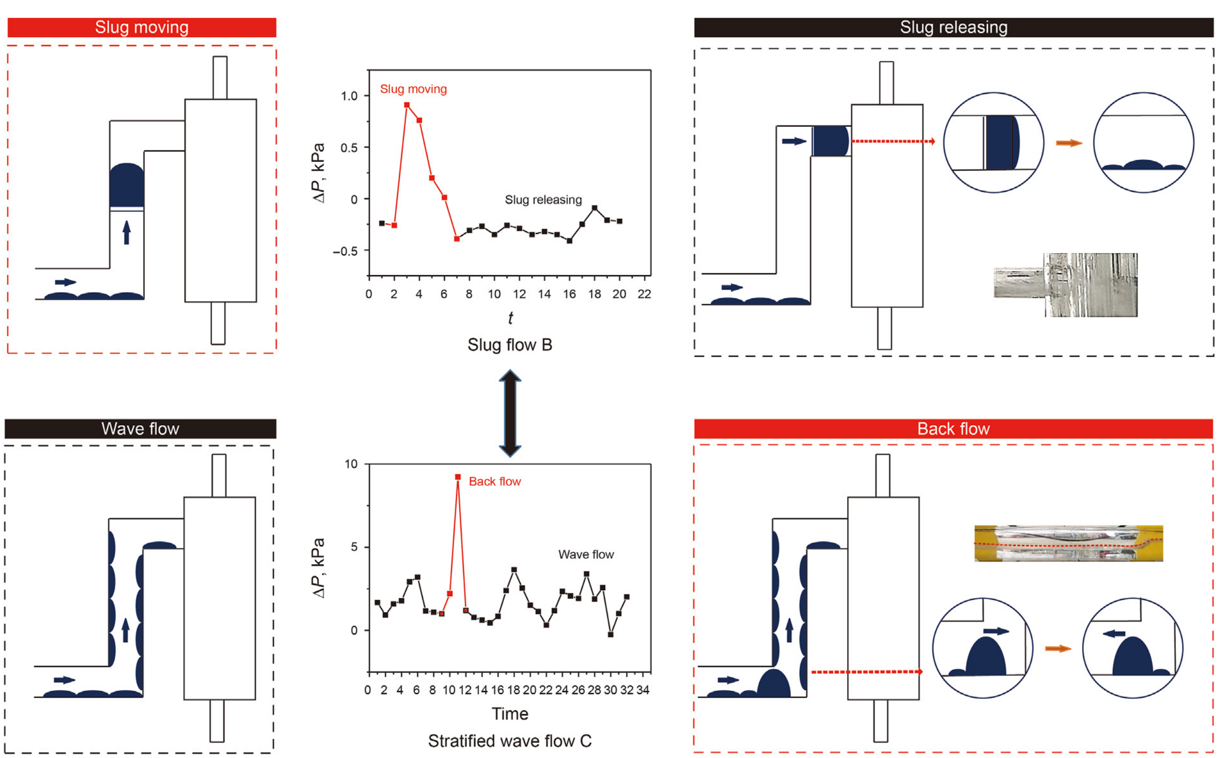

Fig. 17. Example of slug flow A and pressure drop.

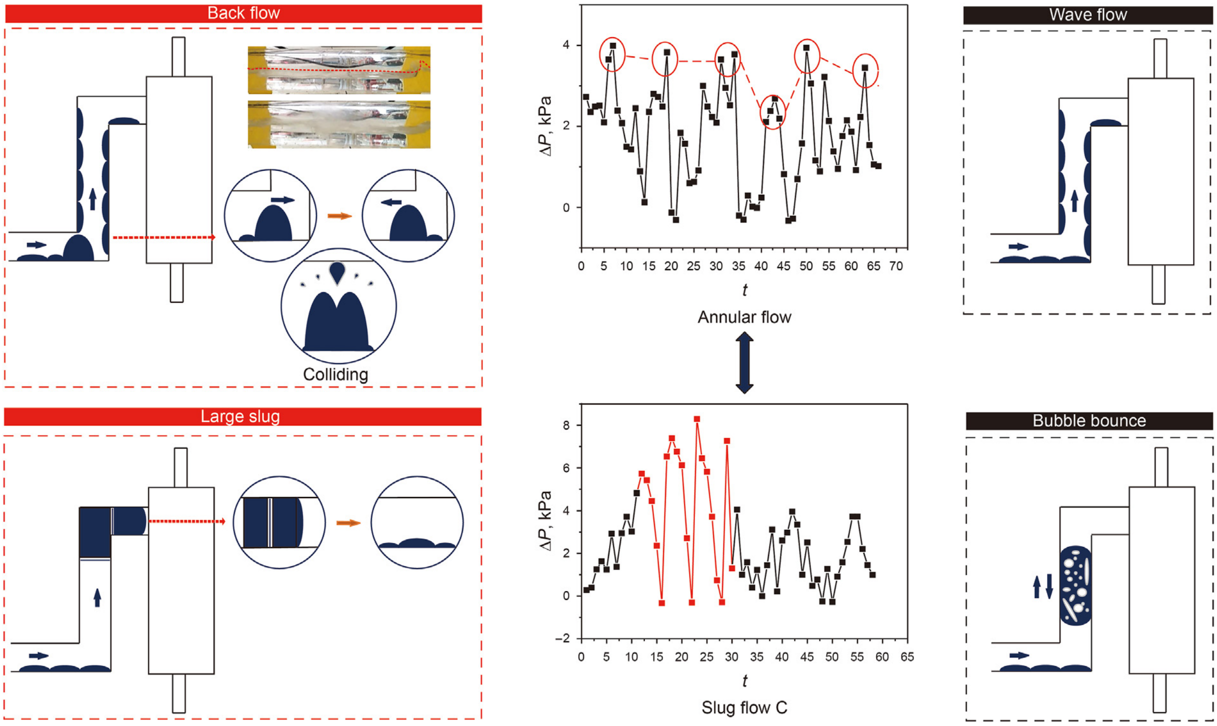

Fig. 18. Example of slug flow C and pressure drop.

in the identification using neither superficial velocity nor pressure drops. In addition, for KNN, Decision tree, LGBM, these four algorithms are obviously accurate in the flow regimes identification by superficial velocity but not by pressure drops, so in the GLCC flow regimes identification, the machine learning algorithms mentioned above are suitable for the static data such as velocities, but not for

Table 6
Conclusions of machine learning algorithms for GLCC flow regimes identification.

Input data category	Suitable/ unsuitable	Algorithm name	Reasons & remarks
Superficial velocity	suitable	Random forest XGBoost GBDT Bagging Tree Decision Tree LGBM	● The principles of algorithms based on decision tree are consistent with the key of flow regimes identifications, which are thresholds setting ● The classification based on decision tree which can identify the flow regimes step by step is effective ● The gas and liquid superficial velocities make the flow regimes identification simple for tree models and they are suitable indexes as the machine learning input data
		MLP	MLP is suitable for various kinds of data due to its multiple layers and nodes to fit the objects relations
		KNN	Based on distance measurement to classify which is effective for two-dimensional points classifications
	unsuitable	SVM	Most suitable for dichotomy problems
		Adaboost	Giving too many weights to the difficult samples but the data are imbalanced
		Bernoulli Bayesian Network Gaussian Bayesian Network	Requiring the characteristic properties of flow regimes data have no relationship which is not in reality
Pressure drop	suitable	GBDT XGBoost	The combinations of decision tree and Boosting tree, especially with the optimizations of sampling methods and loss function, are suitable for the time-varying pressure drops data, to identify the GLCC flow regimes
	unsuitable	Random forest Bagging Tree Decision Tree LGBM	● The complexities of GLCC structures and multiphase flow in it result in the complexities of pressure drops, which are not suitable to be used for flow regimes identification ● The Decision Tree itself and Bagging Tree algorithms are not suitable for the time-varying pressure drops data
		MLP	● The applied MLP algorithm is original version without optimization and improvement, but it still has accuracy of 0.89 ● For the types of time-varying data, MLP can get into extreme value easily ● MLP has complexity in its layers and nodes, and it has large potential to be optimized and improved
		KNN	The distance measurement is not suitable for the pressure drop data
		SVM	Most suitable for dichotomy problems
		Adaboost	Giving too many weights to the difficult samples but the data are imbalanced
		Bernoulli Bayesian Network Gaussian Bayesian Network	Requiring the characteristic properties of flow regimes data have no relationship which is not in reality

the dynamic data such as pressures. However, the improvements of Decision tree in the forms of XGBoost and GBDT are effective and can identify the flow regimes by pressure accurately. Especially for KNN, which realizes the identification by distance measurements, the distance measurement is an effective way to describe the boundary between flow regimes, like drawing the flow regimes map. For the MLP algorithm, since it is a complex and special algorithm, and it can easily get into extreme value easily, this algorithm has the accuracy of 0.99 and 0.89 specifically when using the velocity data and pressure drop data, but the MLP itself has large potential to be improved. The multi-layers inside MLP including the number of nodes and the number of layers, but the improvement of which is still the recognized difficult problem in the whole machine learning area. However, the pressure drops of flow regimes has the similar part such as the little fluctuations when the slug is developing. In reality, it is really hard to say some machine learning are absolutely not suitable for the flow regimes identification, because the algorithms of machine learning can be improved by combining with other algorithms or adjusting the index of the function inside. The machine learning algorithms for GLCC flow regimes identification are concluded in Table 6.

4. Conclusion and future works

The pressure drops and the gas/liquid superficial velocities are set as the input data of traditional machine learning algorithms, to realize the GLCC flow regimes identification. In this paper, the availability of data for the practical fields has been considered. Ensemble algorithms, SVM, KNN, Bayesian Models and MLP, which have been classical and popular algorithms of machine learning for identification, are applied in the GLCC flow regimes identification.

For the superficial velocities as the input data, the results show that Random Forest, XGBoost, GBDT, Bagging Tree, Decision Tree, LGBM and MLP can identify the flow regimes by the velocity data with the accuracy above 0.99. The confusion of these algorithms is from stratified smooth flow and slug flow, which are the flow regimes at the inlet of GLCC. The reasons for the confusion are mainly based on the presence of transition area and the overlap of the input data. It has been found that the ensemble algorithms, which are based on the decision tree model, have the high accuracy in flow regimes identification. After the specific analysis of the Bagging and Boosting algorithms, the calculation process of the labeled velocity and flow regimes data, the data flow and details of the two algorithms are arranged clearly, so their success of flow regimes identification combined Decision Tree is the thresholds setting, which continues to be updated and improved in calculation, and this is the key of the identifications. The superficial velocities are effective index for machine learning to identify the GLCC flow regimes, by which the traditional algorithms can have accuracy above 0.99 and be stable after multi-tests without additional improvement, and based on that, Bagging algorithm is better than Boosting. For the algorithms which are not suitable, they are also analyzed by their principles and inapplicability.

For the pressure drops as the input data, which is changing with time periodically, the XGBoost and GBDT can identify the flow regimes accurately during the multi-test, and the results of algorithms have the fluctuations in tests. It has been found that the confusions of other algorithms are from slug flow and stratified wave flow, and from annular flow and slug flow. These flow regimes are analyzed by checking their different evolution processes, and the slug flow A has a slug moving stage which is similar with the springback stage of stratified wave flow. Meanwhile, the slug flow C has the bubble bounce stage which is similar with the colliding of the inverse flows in horizon pipe from annular flow. Based on the experimental flow regime evolution processes and the corresponding pressure drops, and combined with the analysis of the algorithms principles and the research of GLCC flow regimes before, the success of XGBoost and GBDT, and the confusion of other algorithms are explained. Compared with superficial velocities, the pressure drops are not as suitable as that to identify the GLCC flow regimes, because of the complexity and insufficiency of pressure drops data, and the limitation of the algorithms. For the algorithms which are suitable for the velocity data but not for pressure drop data, their applicability and inadaptability are analyzed based on the calculation processes and model principles. The measurement of gas/liquid flow rate is more difficult than that of pressure drops. It should be furtherly clarified that for the GLCC flow regimes identification based on velocities and pressure drops, in terms of accuracy and algorithms applicability, gas/liquid superficial velocity is better; however, for the practical fields which cannot measure gas and liquid flow rate respectively, the pressure drops should be chosen.

There is a long way to go for the machine learning applied in the GLCC flow regimes identification. The explanation based on physical mechanisms of machine learning is important, because it is an effective way to let practical industry fields accept that, and to check the rationality of the learning results. The accuracies of machine learning are related to the amount of flow regimes data and the description of the transition area between flow regimes. In addition, the indexes adjustments and the optimization of the algorithms are important, so as to be more suitable for the complexity of the multiphase in GLCC and the complexity of the flow regimes types and evolution processes. In view of the industrial application, research should focus more on the GLCC flow regimes identification by data from practical fields.

References

Abellán, J., Masegosa, A.R., 2012. Bagging schemes on the presence of class noise in classification. Expert Syst. Appl. 39, 6827–6837. https://doi.org/10.1016/j.eswa.2012.01.013.

Arpandi, I.A., Joshi, A.R., Shoham, O., Shirazi, S.A., Kouba, G.E., 1996. Hydrodynamics of two-phase flow in gas-liquid cylindrical cyclone separators. SPE J. 1, 427–436. https://doi.org/10.2118/30683-PA.

Chirinos, W.A., Gomez, L.E., Wang, S., Mohan, R.S., Shoham, O., Kouba, G.E., 2000. Liquid carry-over in gas/liquid cylindrical cyclone compact separators. SPE J. 259–267. https://doi.org/10.2118/65094-PA.

Du, M., Yin, H., Chen, X., Wang, X., 2018. Oil-in-water two-phase flow pattern identification from experimental snapshots using convolutional neural network. IEEE Access 7, 6219–6225. https://doi.org/10.1109/ACCESS.2018.2888733.

Gorban, A.N., Kégl, B., Wunsch, D.C., Zinovyev, A.Y., 2008. Principal Manifolds for Data Visualization and Dimension Reduction, Lecture Notes in Computational Science and Engineering. Springer.

Hobold, G.M., da Silva, A.K., 2018. Machine learning classification of boiling regimes with low speed, direct and indirect visualization. Int. J. Heat Mass Tran. 125, 1296–1309. https://doi.org/10.1016/j.ijheatmasstransfer.2018.04.156.

Hreiz, R., Lainé, R., Wu, J., Lemaitre, C., Gentric, C., Fünfschilling, D., 2014. On the effect of the nozzle design on the performances of gas-liquid cylindrical cyclone separators. Int. J. Multiphas. Flow 58, 15–26. https://doi.org/10.1016/j.ijmultiphaseflow.2013.08.006.

Iyer, S., Lang, P., Schoppa, W., Chin, D., Leitko, A., 2010. Subsea processing at parque das conchas (BC-10): taking flow assurance to the next level. In: Proceedings of the Annual Offshore Technology Conference. OnePetro, pp. 528–537. https://doi.org/10.4043/20451-ms.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science (80) 349, 255–260.

Ju, G.T., Littell, H.S., Cook, T.B., Dupre, M., Clausing, K.M., Shumilak, E., Schoppa, W.W., Blizzard, W.A., 2010. Perdido development: subsea and flowline systems. In: Proceedings of the Annual Offshore Technology Conference. OnePetro, pp. 2664–2675. https://doi.org/10.4043/20882-ms.

Julia, J.E., Ozar, B., Jeong, J.J., Hibiki, T., Ishii, M., 2011. Flow regime development analysis in adiabatic upward two-phase flow in a vertical annulus. Int. J. Heat Fluid Flow 32, 164–175. https://doi.org/10.1016/j.ijheatfluidflow.2010.09.003.

Khoi Vu, V., Fantoft, R., Shaw, C., Gruehagen, H., 2009. Comparison of subsea separation systems. In: Offshore Technology Conference. OnePetro. https://doi.org/10.4043/otc-20080-ms.

Kouba, G.E., Shoham, O., Shirazi, S., 1995. Design and performance of gas-liquid cylindrical cyclone separators. In: Proceedings of the BHR Group 7th International Meeting on Multiphase Flow. Cannes, France, pp. 307–327.

Kouba, G., Wang, S., Gomez, L., Mohan, R., Shoham, O., 2006. Review of the state-of-the-art gas/liquid cylindrical cyclone (GLCC) technology - field applications. In: International Oil and Gas Conference and Exhibition in China 2006 - Sustainable Growth for Oil and Gas. OnePetro, pp. 893–901. https://doi.org/10.2118/104256-ms.

Kulkarni, S.J., Shinde, N.L., 2016. Studies and research on cyclone separators: a review. Int. J. Sci. Res. Sci. Technol. 1, 28–31. https://doi.org/10.32628/IJSRST.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Liu, L., Bai, B., 2019a. A mechanistic model for the prediction of swirling annular flow pattern transition. Chem. Eng. Sci. 199, 405–416. https://doi.org/10.1016/j.ces.2019.01.039.

Liu, L., Bai, B., 2019b. Flow regime identification of swirling gas-liquid flow with image processing technique and neural networks. Chem. Eng. Sci. 199, 588–601. https://doi.org/10.1016/j.ces.2019.01.037.

Liu, L., Zhang, J., Liu, S., Wang, K., Gu, H., 2021. Decay law and swirl length of swirling gas-liquid flow in a vertical pipe. Int. J. Multiphas. Flow 137, 103570. https://doi.org/10.1016/j.ijmultiphaseflow.2021.103570.

Manjrekar, O.N., Dudukovic, M.P., 2019. Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique. Chem. Eng. Sci. X 2, 100023. https://doi.org/10.1016/j.cesx.2019.100023.

Massaoudi, M., Refaat, S.S., Chihi, I., Trabelsi, M., Oueslati, F.S., Abu-Rub, H., 2021. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. Energy 214, 118874. https://doi.org/10.1016/j.energy.2020.118874.

Mi, Y., Ishii, M., Tsoukalas, L.H., 1998. Vertical two-phase flow identification using advanced instrumentation and neural networks. Nucl. Eng. Des. 184, 409–420. https://doi.org/10.1016/S0029-5493(98)00212-X.

Mogseth, G., 2008. Functional verification of the worlds first full field subsea separation system - TIORA. In: Offshore Technology Conference. OnePetro. https://doi.org/10.4043/19328-ms.

Moncayo, J.A., Dabirian, R., Mohan, R.S., Shoham, O., Kouba, G., 2018. Foam break-up under swirling flow in inlet cyclone and GLCC. J. Pet. Sci. Eng. 165, 234–242. https://doi.org/10.1016/j.petrol.2018.02.027.

Movafaghian, S., Jaua-Marturet, J.A., Mohan, R.S., Shoham, O., Kouba, G.E., 2000. The effects of geometry, fluid properties and pressure on the hydrodynamics of gas-liquid cylindrical cyclone separators. Int. J. Multiphas. Flow 26, 999–1018. https://doi.org/10.1016/S0301-9322(99)00076-2.

Nnabuife, S.G., Pilario, K.E.S., Lao, L., Cao, Y., Shafiee, M., 2019. Identification of gas-liquid flow regimes using a non-intrusive Doppler ultrasonic sensor and virtual flow regime maps. Flow Meas. Instrum. 68, 101568. https://doi.org/10.1016/j.flowmeasinst.2019.05.002.

Paranjape, S., Chen, S.W., Hibiki, T., Ishii, M., 2011a. Flow regime identification under adiabatic upward two-phase flow in a vertical rod bundle geometry. J. Fluids Eng. Trans. ASME 133. https://doi.org/10.1115/1.4004836.

Paranjape, S., Ritchey, S.N., Garimella, S.V., 2011b. Electrical impedance-based void fraction measurement and flow regime identification in microchannel flows under adiabatic conditions. ASME 2011 Pacific Rim Tech. Conf. Exhib. Packag. Integr. Electron. Photonic Syst. InterPACK 2011 (2), 193–202. https://doi.org/10.1115/IPACK2011-52116.

Paranjape, S., Ritchey, S.N., Garimella, S.V., 2011c. Impedance-based void fraction measurement and flow regime identification in microchannel flows. ASME 2011 Pacific Rim Tech. Conf. Exhib. Packag. Integr. Electron. Photonic Syst. InterPACK 2011 (2), 193–202. https://doi.org/10.1115/IPACK2011-52116.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., 2011. Scikit-learn: machine learning in Python. J. Machine Learn. Res. 12, 2825–2830.

Peixoto, G.A., Ribeiro, G.A.S., Barros, P.R.A., Meira, M.A., Barbosa, T.M., 2005. VASPS prototype in marimba field - workover and re-start. In: SPE Latin American and Caribbean Petroleum Engineering Conference Proceedings. OnePetro. https://doi.org/10.2118/95039-ms.

Rebentrost, P., Mohseni, M., Lloyd, S., 2014. Quantum support vector machine for big data classification. Phys. Rev. Lett. 113, 130503. https://doi.org/10.1103/PhysRevLett.113.130503.

Rosa, E.S., França, F.A., Ribeiro, G.S., 2001. The cyclone gas-liquid separator: operation and mechanistic modeling. J. Pet. Sci. Eng. 32, 87–101. https://doi.org/10.1016/S0920-4105(01)00152-8.

Sawant, P., Schelegel, J., Paranjape, S., Ozar, B., Hibiki, T., Ishii, M., 2008. Flow regime identification in large diameter pipe. Proceedings, ICONE 341–351. https://doi.org/10.1115/ICONE16-48311. International Conference on Nuclear Engineering.

Shaban, H., Tavoularis, S., 2014. Identification of flow regime in vertical upward air-water pipe flow using differential pressure signals and elastic maps. Int. J. Multiphas. Flow 61, 62–72. https://doi.org/10.1016/j.ijmultiphaseflow.2014.01.009.

Shoham, O., Kouba, G., 1998. State of the art of gas/liquid cylindrical-cyclone compact-separator technology. J. Petrol. Technol. 50, 58–65.

Sunde, C., 2005. Classification of two-phase flow regimes via image analysis and a neuro-wavelet approach. Prog. Nucl. Energy 46 (3–4), 348–358. https://doi.org/10.1016/j.pnucene.2005.03.015.

Tambouratzis, T., Pázsit, I., 2009. Non-invasive on-line two-phase flow regime identification employing artificial neural networks. Ann. Nucl. Energy 36, 464–469. https://doi.org/10.1016/j.anucene.2008.12.002.

Tan, C., Dong, F., Wu, M., 2007. Identification of gas/liquid two-phase flow regime through ERT-based measurement and feature extraction. Flow Meas. Instrum. 18, 255–261. https://doi.org/10.1016/j.flowmeasinst.2007.08.003.

Timung, S., Mandal, T.K., 2013. Prediction of flow pattern of gas-liquid flow through circular microchannel using probabilistic neural network. Appl. Soft Comput. J. 13, 1674–1685. https://doi.org/10.1016/j.asoc.2013.01.011.

Wang, H.X., Zhang, L.F., 2009. Identification of two-phase flow regimes based on support vector machine and electrical capacitance tomography. Meas. Sci. Technol. 20, 114007. https://doi.org/10.1088/0957-0233/20/11/114007.

Wang, Y., Kandeal, A.W., Swidan, A., Sharshir, S.W., Abdelaziz, G.B., Halim, M.A., Kabeel, A.E., Yang, N., 2021. Prediction of tubular solar still performance by machine learning integrated with Bayesian optimization algorithm. Appl. Therm. Eng. 184, 116233. https://doi.org/10.1016/j.applthermaleng.2020.116233.

Wang, Y.A., Chen, J.Y., Yue, T., Yang, Y., Han, M.S., Yan, C.Y., 2021. Measurement of thickness and analysis on flow characteristics of upper swirling liquid film in gas-liquid cylindrical cyclone. Exp. Therm. Fluid Sci. 123, 110331. https://doi.org/10.1016/j.expthermflusci.2020.110331.

Xu, C., Chen, J., Dong, T., Jia, Z., Zhang, X., 2018. Measurement and Analysis of Liquid Phase Separation Efficiency of Deep Sea Gas-Liquid Cylindrical Cyclones. China Offshore Platf.

Xu, Z., Yang, X., Chen, B., Zhang, M., Li, Y., 2019. Imaging of flow pattern of gas-oil flows with convolutional neural network. In: IST 2019 - IEEE International Conference on Imaging Systems and Techniques, Proceedings. IEEE, pp. 1–6. https://doi.org/10.1109/IST48021.2019.9010576.

Xu, Q., Zhou, H., Zhu, Y., Cao, Y., Huang, B., Li, W., Guo, L., 2020. Study of identification of global flow regime in a long pipeline transportation system. Powder Technol. 362, 507–516. https://doi.org/10.1016/j.powtec.2019.12.018.

Yang, Z., Wang, J., Zhang, J., He, L., Luo, X., Su, H., He, X., 2022. A statistical method for flow pattern and efficiency prediction using pressure drop in a two-stage gas liquid cylindrical cyclone. J. Nat. Gas Sci. Eng. 99, 104414. https://doi.org/10.1016/j.jngse.2022.104414.

Yue, T., Chen, J., Song, J., Chen, X., Wang, Y., Jia, Z., Xu, R., 2019. Experimental and numerical study of upper swirling liquid film (USLF) among gas-liquid cylindrical cyclones (GLCC). Chem. Eng. J. 358, 806–820. https://doi.org/10.1016/j.cej.2018.10.018.

Zhai, L.S., Bian, P., Han, Y.F., Gao, Z.K., Jin, N. De, 2016. The measurement of gas-liquid two-phase flows in a small diameter pipe using a dual-sensor multi-electrode conductance probe. Meas. Sci. Technol. 27, 45101. https://doi.org/10.1088/0957-0233/27/4/045101.