International Journal of Innovative Computing, Information and Control
Volume **12**, Number **6**, December **2016**

ICIC International ©2016 ISSN 1349-4198
pp. 1777–1789

MINE'S PUMP STATION ENERGY CONSUMPTION AND UNDERGROUND WATER DAM LEVELS MONITORING SYSTEM USING MACHINE LEARNING CLASSIFIERS AND MUTUAL INFORMATION ENSEMBLE TECHNIQUE

Ali N Hasan and Bhekisipho Twala

Department of Electrical and Electronic Engineering Technology
University of Johannesburg
P. O. Box 17011, Doornfontein, 2028, Johannesburg, South Africa
{ alin; btwala }@uj.ac.za

Received April 2016; revised August 2016

Abstract. *This paper proposes the use of artificial intelligence techniques for monitoring and predicting the underground water dam level in a double pump station gold mine. Six single classifiers methods (support vector machine, artificial neural network, naïve Bayesian classifier, decision trees, radial basis function and k nearest neighbors) are applied for this purpose. The paper further proposes a new approach to select the most suitable classifiers when constructing the most accurate ensemble in multiple classifier learning. This approach is based on determining the mutual information amount between classifier pairs and is further used to determine the classifiers optimum number in order to build the most accurate ensemble. Simulation results using underground pump station dam levels and energy consumption data show the proposed strategy as being more superior to methods such as Bagging and Boosting techniques in terms of predictive accuracy.*
Keywords: Machine learning classifiers, Ensemble, Mutual information, Energy, Gold mines water pumps

1. **Introduction.** Demand side management (DSM) initiatives were implemented in the South African mining sector due to the shortage of the electricity supply. Energy control systems were installed for specific mining features such as water pumping control systems. These control systems depend on instantaneous data to make a decision and therefore they are not adaptive nor sustainable and needed to be updated frequently in order to meet any alterations in the water pumping system [1,2]. Furthermore, these water pumping control systems are not able to predict the energy consumption or the underground water dam levels. As such, adaptive and predictive monitoring system that could control and predict the energy consumption of the underground water pumping system and dam levels should be introduced. These adaptive and predictive control systems perform based on present and historical data [3].

In deep gold mines, fresh water pumping systems are considered as a main part of gold mines structure. The mine water is basically pumped into the mining levels to reduce its temperature during mining operations. Underground dam levels need to be always monitored and controlled to avoid water flooding or other harm can be caused by low water levels in the dams. Installing intelligent and adaptive control systems for water pump stations could result in increased energy and cost savings as water pump stations consume a large amount of electricity [2,4]. Intelligent control systems use the state of art machine learning (ML) and artificial intelligence techniques. The use of ML in controlling and monitoring water pumps stations will decrease human involvement and possible errors

[5]. Presently, machine learning and artificial intelligence are being applied for prediction, classification and optimization purposes in fields such as robotics, and engineering [5,6].

The major contributions of this paper are:

1)- The introduction and usage of artificial intelligence techniques to control and monitor the water pumping system which could result in a major energy savings in the mines, and that will have a positive impact on the South African national grid, by decreasing the electricity load. On the other hand the adaptive and predictive nature of artificial intelligence techniques will increase the safety of the mine personnel by reducing human errors.

2)- The proposal of a new multiple-classifier learning (ensemble) technique using mutual information theory as means of improving the predictive accuracy for underground dam levels.

This paper is structured as follows: Section 2 gives a background about ensemble techniques. In Section 3, a layout for a mine situated in South Africa is presented. A mutual information theory is presented is Section 4. In Section 5, methods used in our experiments for the paper are described briefly. Comparative experiments on dam levels and energy consumption databases using ensembles and single classifiers are presented in Section 6, followed by the results in Section 5. Section 7 contains concluding remarks.

2. **Ensemble Classifier (Multiple Classifiers).** Ensembles or multi-classifier methods have been recently developed and widely used as learning techniques. Ensembles can be directly implemented, and proven to have an outstanding predictive and classification performance on real-world problems [6].

An ensemble comprises a set of individually trained classifiers (for example neural networks, naïve Bayesian classifier, or both) whose predictions or classifications are joined when classifying distinctive instances. The aim of an ensemble technique is to enhance the prediction results of a statistical learning system. An ensemble classifier is a technique that combines several weak classifiers in order to yield a strong classifier. Many prediction combination techniques of multiple classifiers have been tested and used to produce an ensemble that has a high predictive accuracy and outperforms the single classifier [7]. Ensemble is also known in many other names such as, ensemble methods, committee, classifier fusion, combination, and aggregation [8]. Recent research has illustrated and proved that a strong and good predictive ensemble is the one where the own individual classifiers are the most diverse. In other words, a good ensemble is the one that combines classifiers which make their errors on different parts of the input space [7]. Ensemble techniques usually produce better results and over perform single classifiers when there is a significant diversity and un-correlation among the used methods and algorithms. Therefore, ensembles aim to encourage diversity between the classifiers and techniques they combine [10].

3. **The Mine A Water Pumping System Overview.** Our case study is a South African gold mine, Mine A, which contains two underground water pump stations. There is another adjacent mine that pumps its water to the case study mine through an underground channel connecting the two mines, then the water gets pumped up to the surface. The control system could not adapt to the sudden large amount of water caused by increasing water influx from the adjacent mine, and therefore the control philosophy had to be changed where the mine's water circulation was operated and monitored by the control room operators. The clear-water pumping system is shown in Figure 1.

Figure 1 shows that the mine includes two main pump stations situated in two different mining levels. One pump station in located at 27-level and has 5 water pumps. Every

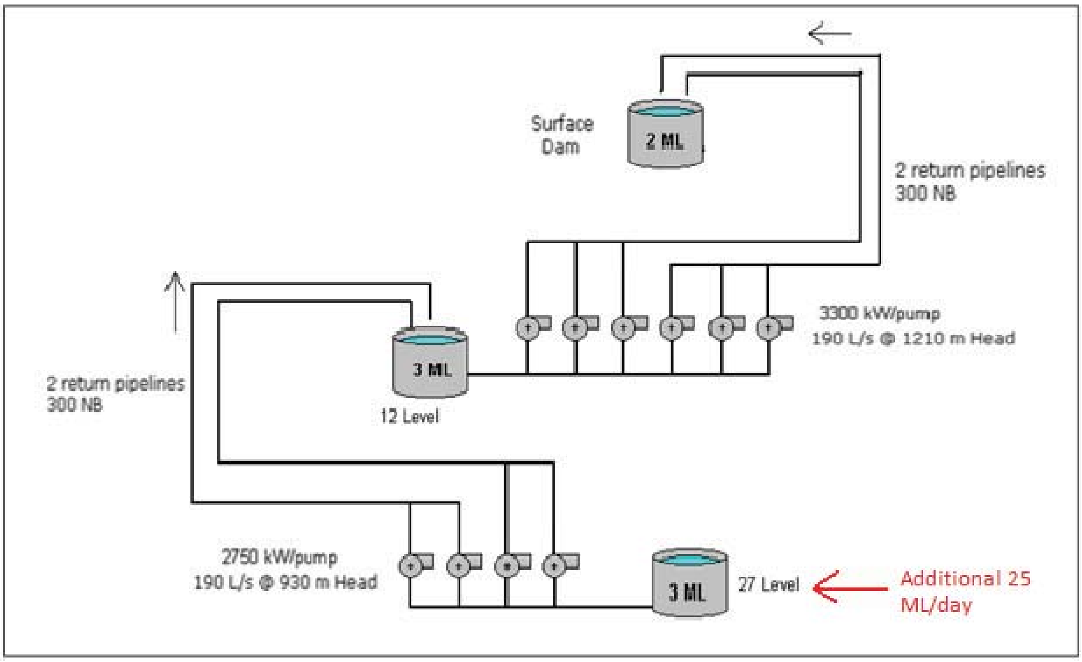

FIGURE 1. Shaft clear-water pumping system

pump has a power capacity of 2.75 MW and a water flow rate of 190 l/s. The total underground dam capacity is 3 Ml. The second pump station is located at 12-level and includes 7 pumps. Each pump is rated at 3.30 MW and can pump 190 l/s. The total underground dam capacity is 2 Ml.

As mentioned before it was essential to frequently update the pump station control system due to the fluctuations of the water volume that may increase or decline as a result of the additional water as shown in Figure 1. Therefore, a machine learning control system is proposed to predict the water levels, henceforward control the water dam levels within the safe limits. This predictive control system uses parameters such as the water levels, the pump running status (on/off), the amount of consumed energy, as explained in Section 6.

4. **Mutual Information Theory.** In this paper, mutual information theory was used to construct the most accurate ensemble. This was achieved by determining the mutual information amount between all used single classifiers. The amount of mutual information indicates the diversity and dependency level of the single classifiers [12].

Mutual information is defined as the measure of how much one random variable is similar to another random variable. Mutual information is usually measured in bits. However, the unit of information depends on the base of the logarithm. When the base is 2, then, the unit of quantity of mutual information is the "bit". When two parties (statistically) have one bit of information in common, then the mutual information quantity in bits is 1. "Entropy" is the important concept of information theory which represents the amount of uncertainty of a given system. It also measures the information quantity that is required usually to define the random variable [12,13]. If the random variable is definitely to be predicted then the entropy value is positive or it is zero. The main objective of mutual information theory is to reduce the entropy amount in a system. If the shared information is too small and insignificant between two items, then these two items are considered to be independent. On the other hand, if the items share big amount of information between each other, then these items are not likely to be independent system; in other words the two items depend one each other [13,14]. Mutual information can be defined as:

$$I(X;Y) = H(X) - H(X/Y)$$

where H(X) is the entropy and H(X/Y) is the conditional entropy. The mutual information is the reduction in the uncertainty (entropy) of X due to the knowledge of Y where X and Y are discrete random variables. Figure 2 shows the relationship between mutual information I(X;Y) and the entropies of two random variables and the conditional entropies [13,14].

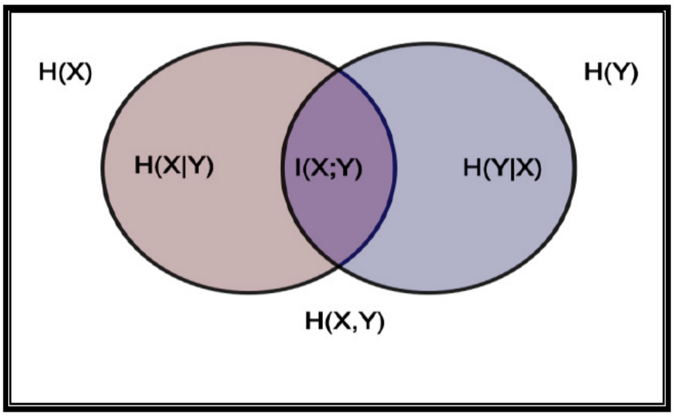

FIGURE 2. The relationship between the entropies of two random variables X and Y and their mutual information

5. **Machine Learning Algorithms Utilized in This Study.** In this study, six classifiers were used in order to monitor and predict the water levels and the energy consumption for an underground pump station. These classifiers have been trained and tested using real time data which was collected from the mine. Default parameters were used for each classifier. In this section a brief overview about each classifier is presented below.

***A. Multilayer Perceptron (MLP)*:** Multilayer perceptron (MLP) is a type of artificial neural network. It computes using a set of weighted connected neurons [15,16]. MLP basic structure usually comprises three layers, an input layer, hidden layer and an output layer. A layer consists of a set of neurons, and each neuron utilizes a nonlinear activation function. Feed-forward setup is used to connect the neurons. Each neuron in a layer connects with a certain weight to every neuron in the next layer [16,23]. MLP is considered among the most powerful techniques in solving difficult and different problems by training them in a supervised mode with a famous algorithm known as the back propagation algorithm. Learning in the perceptron is accomplished by modifying connection weights after all the data cases are processed, depending on the error volume in the output compared to the anticipated outcome [15,16].

***B. Support Vector Machine (SVM)*:** In machine learning and pattern recognition, support vector machine classifier (SVM) is identified as a superb classifier [17,18]. SVM is also proven to be powerful when used to resolve two-class (binary) classification tasks. The mechanism that SVM classifier follows is to determine a linear maximum margin hyper-plane within the spaces of the data-points that deliver the utmost distance amid the two classes. The closest data-points to the linear maximum margin hyper-plane from the support vectors are considered classified correctly [17,18,23].

***C. Decision Trees (DT)*:** A decision tree classifier (DT) is a classifier that utilizes the tree-like model to make decisions and reach their possible consequences. It is one system to show an algorithm [19]. Decision tree forms an interpretable model that denotes a set of rubrics. It is a famous and well-known technique for classification. DT is relatively fast method to train and to produce predictions. Decision trees are applied into many fields, such as operations research, especially in decision analysis, to identify a strategy most likely to reach a goal [19,20].

***D. Naïve Bayes' Classifier (NBC)*:** The naïve Bayes' classifier gives a simple approach, with clear semantics, to representing, using, and learning probabilistic knowledge [21]. Basically, a naïve Bayes' classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable [22]. It is based on applying Bayes' theorem with strong (naïve) independence assumptions, or more specifically, independent feature model [22].

***E. Radial Basis Function Classifier (RBF)*:** The radial basis function (RBF) network is a special type of neural networks with several distinctive features [24]. An RBF network is structured in a way that it has a feed-forward neural network as an input layer, then there is a hidden layer which contains the processing units that uses RB function [25]. Each unit in the hidden layer then produces an activation based on the associated radial basis function. Lastly, there is an output layer that computes a linear combination of the activations of the hidden units [26].

***F. K Nearest Neighbors (k-NN)*:** In machine learning, k-NN represents a memory-based (or data-point-based) technique for classification and regression problems. k-NN is a popular classification technique that is successfully implemented in many applications such as medical research, and business applications. The aim of the k-nearest neighbors (k-NN) method is to use a data-set in a way the data points are classified into few different classes for the prediction and classification of a new sample point [22]. In classification problems, the label of possible items is obtained using the labels of closest training data points in the feature space. This item is obtained by usually using "majority voting" or "averaging" techniques [27].

6. Simulation Models and Experiments.

6.1. Experiment one set-up.

The first experiment was to train and test each single classifier separately from the six classifiers group (MLP, k-NN, DT, NBC, RBF, and SVM) using the mine pumps energy data and underground dam levels. Data for energy consumption and underground dam levels was collected for a period of three months by using pressure transmitter fitted on the dams. This pressure transmitter is connected to a programmable logic controller (PLC) fixed on the pump station, then via fibre optics to a supervisory control and data acquisition (SCADA) system to log the data on a spread sheet. It records a value every two seconds. Figure 3 illustrates the proposed machine learning (ML), monitoring, and control system, and it also shows the required infrastructure such as the PLC's on each pump station and the SCADA PC which is connected to the ML server. The energy consumption data represents the run time status for each pump, where 1 value denotes pump is on status, and 0 is off status. Each pump has a specified power capacity, as mentioned before, which determines the amount of power consumption.

All the six machine learning algorithms used default parameters employed. For simulation purposes, the data was averaged over 30 second intervals. Starting with the pumps energy data for 12-level and 27-level pumps ML classifiers were trained and tested. Energy data was categorized into classes for 12-level as shown in Table 1. The energy data represent the energy consumed by the underground pumps for 12-level pump station.

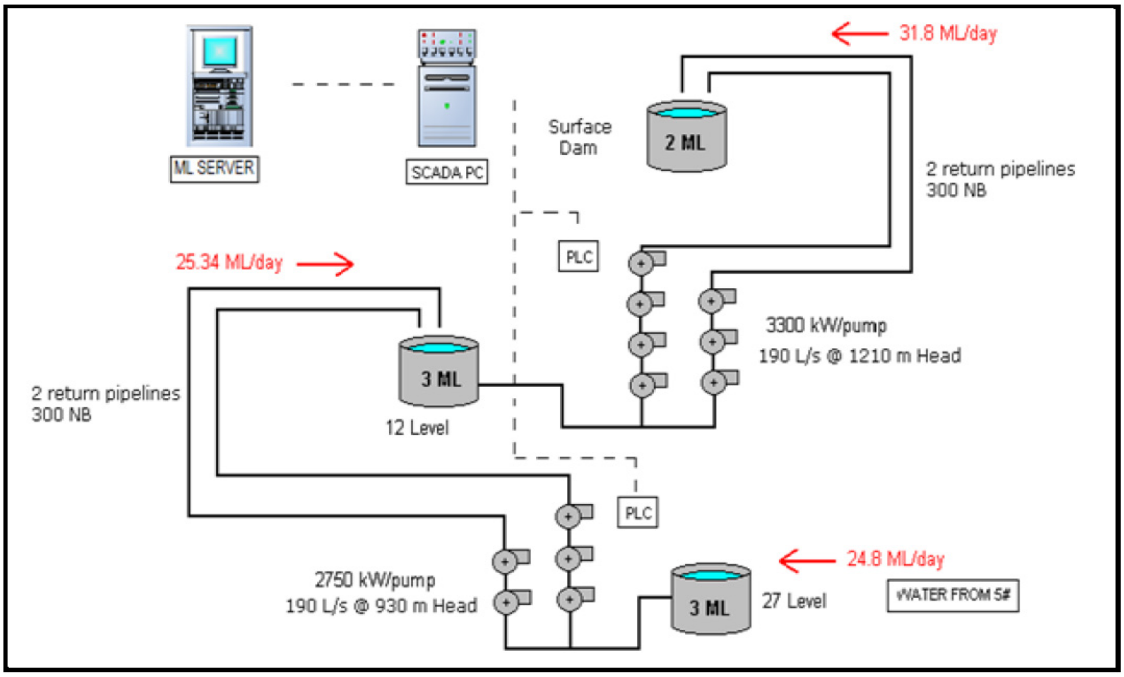

Figure 3. ML monitoring and control system

Table 1. 12-level energy consumption classes

Energy level (kW)	Description	Class
0	No Energy Consumption	0
3300	Low Energy Consumption	1
6600	Medium Energy Consumption	2
9900	Medium Energy Consumption	3
13200	Relatively High Energy Consumption	4
16500	High Energy Consumption	5
More than 16500	Very High Energy Consumption	6

Table 2. 27-level energy consumption classes

Energy level (kW)	Description	Class
0	No Energy Consumption	0
2750	Low Energy Consumption	1
5500	Medium Energy Consumption	2
8250	Relatively High Energy Consumption	3
11000	High Energy Consumption	4
More than 11000	Very High Energy Consumption	5

For 27-level underground pump station, energy data are categorized for classes as shown in Table 2. The energy data represent the energy consumed by the underground pumps. The data was split into 80% for training, and 20% for testing intervals for both pumping stations. As mentioned before, the pumps are directly linked to the dam level, so the attributes here represent the pump running status (on, off) in the case of energy consumption test. The class represents, the total energy consumed by the pumps.

For underground dam level, each level (12-level and 27-level) data was trained and tested separately. The maximum and minimum dam levels for both levels in this mine

Table 3. 27-level dam level classes

Description	Dam level percentage	Class
Pump damage risk	> 25%	1
Low	25%-39%	2
Medium	40%-67%	3
High	68%-85%	4
Critically high (flooding risk)	> 85%	5

Table 4. 12-level dam level classes

Description	Dam level percentage	Class
Pump damage risk	< 24%	1
Critically low	24%-30%	2
Low water level	31%-42%	3
Medium water level	43%-78%	4
High water level	79%-85%	5

were given by the mine's engineer. For 27-level the maximum is 85% and the minimum is 20%, and the same for 12-level underground dam. The data was categorized into classes for simulation. Classes for 27-level dam levels are shown in Table 3.

Table 4 shows 12-level underground dam level percentages and the corresponding classes. Each dam was classified to 5 classes which represent a certain water level percentage. These classes were carefully chosen to suit the underground dam capacity and to precisely monitor and control the dams within the required safe limits.

6.2. **Experiment one results.**

A. Pumps energy consumption

For the pumps energy consumption experiment, the number of instances for energy data is 5, 620,200. After identifying the required classes, the data was processed by WEKA software to define the most accurate classifier(s) that will achieve the maximum prediction accuracy between the six classifiers used. Table 5 shows the classifiers prediction accuracies and errors for 12-level pumps energy consumption.

Table 5. 12-level pumps energy consumption prediction results

Description	MLP	SVM	DT	NBC	*k*-NN	RBF
Misclassification error	**1.34%**	1.071%	9.396%	12.080%	6.040%	1.60%
Mean absolute error	**0.0044**	0.1090	0.0160	0.0445	0.0087	0.0082
Root mean squared error	**0.040**	0.2080	0.2993	0.1220	0.0808	0.6300

It can be seen from Table 5 that SVM has outperformed the other classifiers and achieved the highest prediction accuracy with nearly 99%. In terms of the mean absolute error and root mean square error, MLP performed the best; therefore, it may be considered to be the best classifier on these two bases. RBF and k-NN achieved excellent results followed by DT. According to "Tukey-multiple test" the performance of these three classifiers (SVM, MLP, and RBF) is almost the same. NBC achieved the least predictive accuracy and the highest misclassification error of 12.08%, followed by DT with 9.396% misclassification accuracy with significance difference in performance.

For 27-level pump station energy consumption, MLP was the best classifier in terms of all performance measures and achieved 99% prediction accuracy (Table 6). RBF and

Table 6. 27-level pumps energy consumption prediction results

Description	MLP	SVM	DT	NBC	*k*-NN	RBF
Misclassification error	**1.00%**	2.00%	11.456%	11.150%	3.067%	1.49%
Mean absolute error	**0.0053**	0.1130	0.1260	0.0993	0.0068	0.0082
Root mean squared error	**0.033**	0.2370	0.1088	0.1310	0.0710	0.6410

k-NN classifiers showed an acceptable predictive performance, while DT and NBC underperformed the six classifiers in terms of all performance measures. After applying "Tukey-multiple test" for these predictive results, significance difference appeared in performance between MLP, NBC and DT classifiers.

B. Underground dam level's results

Prediction results for 12-level pump station underground dams are shown in Table 7.

Table 7. 12-level underground dam levels prediction results

Description	MLP	SVM	DT	NBC	*k*-NN	RBF
Misclassification error	**46.30%**	45.60%	48.40%	47.00%	46.67%	46.89%
Mean absolute error	**0.179**	0.216	0.189	0.183	0.207	0.184
Root mean squared error	**0.306**	0.320	0.318	0.309	0.307	0.312

Table 7 shows that all six methods have almost the same predictive accuracy, with SVM being slightly more accurate, with minimal differences in root mean squared error and mean absolute error. However, in terms of all measures, MLP shows the best performance with the lowest mean absolute error and root mean square error as there is no significance difference in performance between SVM and MLP. In general, all six classifiers demonstrated low prediction accuracy.

Results for 27-level underground dam level prediction were similar to the prediction results for 12-level underground dams in terms of high misclassification error and low accuracy (Table 8). In terms of all performance measures NBC was the best performer.

Table 8. 27-level underground dam levels prediction results

Description	MLP	SVM	DT	NBC	*k*-NN	RBF
Misclassification error	**42.95%**	**42.95%**	51.40%	43.10%	43.67%	**42.95%**
Mean absolute error	**0.232**	0.2695	0.239	0.212	0.246	0.241
Root mean squared error	0.346	0.360	0.344	0.341	0.356	0.349

6.3. **Experiment two set-up.** It can be seen from the single classifier results that the classifiers achieved high prediction accuracy for energy data but did not perform as well for the underground dam levels, which necessitated the implementation of an ensemble. The second experiment is to combine the classifiers using multi-classifiers (ensemble) technique in order to investigate the possibility of improving the prediction accuracy of the underground dam levels.

Our first approach for combining the classifiers was to use Bagging and Adaboost ensemble techniques. Different splits of train-to-test percentages were tested and the most accurate split was found to be 60% to train and 40% to test for the Bagging and Adaboost ensemble experiments. WEKA software was used to simulate the experiment.

Our second attempt entailed the construction of an ensemble depending on the mutual information amount between the classifiers using a MATLAB simulator that was programmed for this purpose. The underground water dam level's data was organized in arrays as an ".m" file which can deal with MATLAB function, script, or class to be suitable for the MATLAB simulator. All the six classifiers were included in the simulator and use majority vote algorithm to combine the classifiers output. Majority vote was used since it gives the best results among other combination algorithms, such as the sum, the maximum, the minimum, the average, products and the Bayes algorithms for this particular case.

The initial assumption states that, increasing diversity may decrease ensemble error. The ensemble was constructed based on the amount of mutual information between each two classifiers to determine the diversity between the classifiers. Each two classifiers predicted classes' vectors mutual information value was determined using the water dam levels dataset. The lowest shared information classifiers were considered the most independent, thus the best classifiers to build the ensemble. The ensemble was only used for the underground water dam's level as will be seen later. The steps below summarize the ensemble construction.

1)- Divide original dataset into training datasets, and testing dataset for all classifiers.
2)- Calculate mutual information between the six classifiers by taking every two classifiers together, hence calculating 15 mutual information values.
3)- Combine the classifiers starting with the lowest mutual information (the most independent classifiers) to construct the first ensemble.
4)- Add the second lowest mutual information classifier to the first ensemble to construct ensemble two.
5)- Construct ensemble three and ensemble four by adding the remaining classifiers.
6)- Combine the multiple classifiers into aggregate output using majority voting algorithm.
7)- Compare the classification accuracy of the various ensembles.

6.4. **Experiment two results.** Firstly, Bagging and Adaboost ensemble methods were experimented and simulated for 12-level underground water dam levels data. Prediction accuracy was (surprisingly) not increased compared to the single classifiers prediction results with Bagging and Adaboost both achieved approximately **53%** accuracy. For 27-level underground water dam levels Bagging and Adaboost achieved the same prediction accuracy of **57%**; hence there was no improvement in terms of the predictive accuracy when compared to the single classifiers performance.

For the mutual information ensemble experiment, the optimum number of classifiers that construct the ensemble were determined based on the amount of shared information between the six classifiers. This was done by taking two classifiers at a time and compute the mutual information value which indicates the most independent classifiers, hence the best to build the most accurate ensemble.

Starting with 12-level underground dam levels ensemble structures, 15 pairs of classifiers were considered; hence, 15 values of mutual information were computed for each level. Table 9 shows the mutual information values for 12-level pump station data. It can be seen that the lowest mutual information achieved correspond to the pairs (MLP, k-NN), (MLP, SVM) and (MLP, NBC) with value of 0.2032 bit, which means that MLP, k-NN and NBC are the most independent classifiers. (MLP, SVM) also has low mutual information compared to (MLP, DT). It can also be noticed that DT has the highest mutual information with all the 5 classifiers. The highest mutual information amount was recorded between the classifier pair (NBC, DT) with 1.7830 bit.

From the mutual information results, the first ensemble (ens1) will comprise MLP and k-NN. The second ensemble (ens2) will combine MLP, k-NN and NBC. SVM will be

TABLE 9. 12-level classifiers mutual information

Number	Classifiers pairs	Mutual information (bit)
1	MLP, k-NN	**0.2032**
2	MLP, SVM	**0.2032**
3	MLP, RBF	0.4566
4	MLP, NBC	**0.2032**
5	MLP, DT	1.0582
6	k-NN, SVM	**0.2313**
7	k-NN, RBF	0.4076
8	k-NN, NBC	**0.2345**
9	k-NN, DT	0.8695
10	SVM, RBF	1.3903
11	SVM, NBC	0.3795
12	SVM, DT	0.8991
13	RBF, NBC	0.6193
14	RBF, DT	0.8996
15	NBC, DT	1.7830

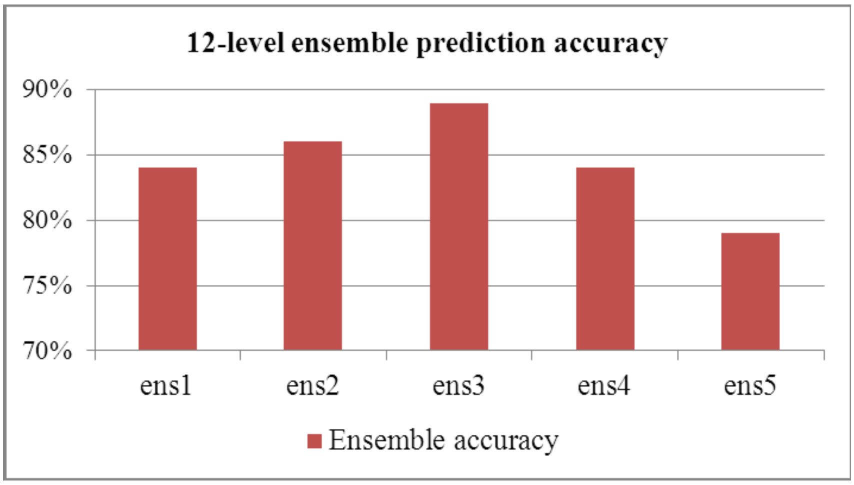

FIGURE 4. 12-level mutual information ensemble results

added to ensemble two to construct the third ensemble (ens3). The fourth ensemble (ens4) was constructed by adding RBF to ens3, and the last ensemble "ens5" was built by the addition of DT method to ensemble 4; therefore, ensemble 5 contained all 6 classifiers. The five mutual information ensembles were implemented as follows:

- ens1: MLP + k-NN
- ens2: MLP + NBC + k-NN
- ens3: MLP + NBC + k-NN + SVM
- ens4: MLP + SVM + k-NN + NBC + RBF
- ens5: MLP + SVM + k-NN + NBC + RBF + DT

The five mutual information ensembles were trained and tested using the underground dam levels real-time data. The main performance measure used for the ensemble is the classification accuracy. The ensemble prediction results are shown in Figure 4.

It can be seen that ensemble 3 was the most accurate ensemble and reached a prediction accuracy of 89% with significance difference in performance when compared to the other four ensembles used. Ensemble 2 and ensemble 4 achieved 86% and 84% prediction accuracy respectively with significant difference in performance. Ensemble 1 which consists of MLP and k-NN realized 84% of prediction accuracy. The fifth ensemble (ens5) which contains DT achieved the lowest prediction accuracy of 79% with significance difference in performance. Table 10 shows the mutual information values between the six classifiers for the 27-level underground dam levels data.

TABLE 10. 27-level classifiers mutual information

Number	Classifiers pairs	Mutual information (bit)
1	MLP, k-NN	**0.2011**
2	MLP, SVM	**0.2011**
3	MLP, RBF	0.5061
4	MLP, NBC	**0.2011**
5	MLP, DT	0.9289
6	k-NN, SVM	**0.3003**
7	k-NN, RBF	0.4786
8	k-NN, NBC	**0.2349**
9	k-NN, DT	0.9605
10	SVM, RBF	1.0923
11	SVM, NBC	0.3491
12	SVM, DT	0.9901
13	RBF, NBC	0.709
14	RBF, DT	1.1091
15	NBC, DT	1.8231

It can be seen from Table 10 that the classifier pairs, (MLP, NBC), (MLP, SVM) and (MLP, k-NN) has the lowest shared information with 0.2011 bit, followed by the classifier pair (k-NN, NBC) with mutual information value of 0.2349 bit. The classifier pair (k-NN, SVM) has a relatively low mutual information amount of 0.3003 bit. The highest mutual information value was recorded for the classifier pair (NBC, DT). In general, classifiers pairs containing DT have the highest mutual information amount. Obtaining the mutual information amount made it possible to identify the most diverse and independent classifiers, which lead to building the most accurate ensemble. Figure 5 shows the performance results for the five multi-classifiers (ensembles).

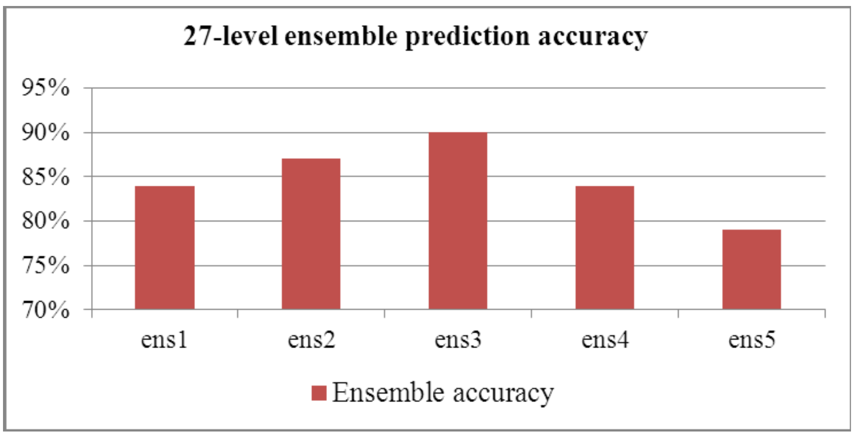

FIGURE 5. 27-level mutual information ensemble results

Figure 5 shows (ens3) with highest predictive accuracy of 90%. (ens2) with 87% predictive accuracy. (ens5) which includes DT underperformed the rest of the five ensembles.

7. **Conclusions and Discussions.** This paper introduced a machine learning based monitoring system which uses single and combined classifiers (ensembles) for a gold mine clear-water pumping system. Two experiments were conducted over two pump stations situated in 12-level and 27-level mining levels in terms of predicting energy consumption and underground dam levels.

The first experiment was to train and test six classifiers (MLP, k-NN, SVM, DT, NBC and RBF) separately using a real-time pumps energy and dam levels data that was collected over 3 months. The results for both pump stations energy consumption's experiment were good as classifiers achieved high prediction accuracy. The DT classifier achieved the lowest prediction accuracy. However, MLP was the best classifier in terms of all performance measures. For the underground dam levels single classifiers experiment, the classifiers prediction performance did not match the energy prediction performance. However, DT was the worst performers among the used classifiers.

Mutual information based ensemble technique was implemented due to the "relatively" poor predictive performance of the single classifiers, Bagging and Adaboost ensemble methods for the dam levels prediction accuracies. The mutual information amount between the single classifiers was computed, and the two classifiers with the least shared amount are considered the most diverse and hence the best to construct the ensemble. Five ensembles were built based on the amount of mutual information.

12-level underground dam level (ens3) which consists of classifiers "MLP, NBC, k-NN, and SVM" was the most accurate ensemble and achieved 89% prediction accuracy. The least accurate ensemble was (ens5) which was constructed from all six classifiers and achieved 79% prediction accuracy. 27-level underground dam level (ens3) which consists of classifiers "MLP, k-NN, SVM, and NBC" was the best ensemble in terms of prediction accuracy with 90% prediction accuracy, while the poorest ensemble prediction performance was recorded by (ens5) which consists of all six classifiers.

The experimental results demonstrate that using artificial intelligence in certain aspects in the mining sector could be possible and might lead to improved safety of the mine personnel and providing enhanced protection to the equipment from damage. Also, energy savings could be realized by regulating the pumps operation schedule.

Finally, the novel uses of new ensemble technique by defining the mutual information amount between classifiers exposed a success in defining the optimum classifiers numbers which construct the most accurate ensemble and displayed encouraging improvement in terms of prediction accuracy of underground dam's water level.

REFERENCES

[1] A. N. Hasan, B. Twala and T. Marwala, Moving towards accurate monitoring and prediction of gold mine underground dam levels, *International Joint Conference on Neural Networks, IEEE World Congress on Computational Intelligence*, Beijing, China, 2014.

[2] A. N. Hasan, K. Ouahada, T. Marwala and B. Twala, Optimization of the compressed air-usage in South African mines, *Africon IEEE*, Livingstone, Zambia, 2011.

[3] J. I. G. Bredenkamp, L. F. van der Zee and J. F. van Rensburg, Reconfiguring mining compressed air networks for cost savings, *IEEE International Conference on the 11th Industrial and Commercial Use of Energy*, South Africa, 2014.

[4] A. N. Hasan, B. Twala and T. Marwala, Predicting mine dam levels and energy consumption using artificial intelligence methods, *IEEE Symposium Series on Computational Intelligence*, Singapore, 2013.

[5] I. K. Kapageridis, Artificial neural network technology in mining and environmental applications, *Mine Planning and Equipment Selection*, Department of Geotechnology and Environmental Engineering, Technological Education Institute of West Macedonia, Greece, 2002.

[6] A. N. Hasan, B. Twala and T. Marwala, Gold mine dam levels and energy consumption classification using artificial intelligence methods, *IEEE Business Engineering and Industrial Applications Colloquium*, Malaysia, 2013.

[7] Y. Bi, J. Guan and D. Bell, The combination of multiple classifiers using an evidential reasoning approach, *Artificial Intelligence*, vol.172, no.15, pp.1731-1751, 2008.

[8] D. Opitz and R. Maclin, Popular ensemble methods: An empirical study, *Journal of Artificial Intelligence Research*, vol.11, pp.169-198, 1999.

[9] Q. Sun and B. Pfahringer, *Bagging Ensemble Selection*, University of Waikato, Hamilton, New Zealand, 2010.

[10] P. Buhlmann, Bagging, boosting and ensemble methods, *ETH Zurich, Seminar fur Statistik, HG G17, CH-8092*, Zurich, Switzerland, 2010.

[11] Y. Rena, P. N. Suganthana and N. Srikanthb, Ensemble methods for wind and solar power forecasting – A state-of-the-art review, *Renewable and Sustainable Energy Reviews*, vol.50, pp.82-91, 2015.

[12] L. de Campos, A scoring function for learning Bayesian networks based on mutual information and conditional independence tests, *The Journal of Machine Learning Research*, vol.7, pp.2149-2187, 2006.

[13] C. D. Manning, P. Raghavan and H. Schütze, *An Introduction to Information Retrieval*, Cambridge University Press, 2008.

[14] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Version 7 (3rd Printing), Cambridge University Press, Cambridge, 2003.

[15] H. C. Peng, F. Long and C. Ding, Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.8, pp.1226-1238, 2005.

[16] A. Taravat, S. Proud, S. Peronaci, F. Del Frate and N. Oppelt, Multilayer perceptron neural networks model for Meteosat second generation SEVIRI daytime cloud masking, *Remote Sensing*, vol.7, no.2, pp.1529-1539, 2015.

[17] A. Folleco, T. Khoshgoftaar, J. van Hulse and A. Napolitano, Identifying learners robust to low quality data, *IEEE International Conference on Information Reuse & Integration*, vol.33, no.3, pp.190-195, 2008.

[18] C. J. C. Burges and B. Schölkopf, Improving the accuracy and speed of support vector machines, *Advances in Neural Information Processing Systems*, vol.9, pp.375-381, 1997.

[19] J. P. Lewis, *A Short SVM (Support Vector Machine) Tutorial*, CGIT Lab/IMSC, Version 0.zz, University of Southern California, 2004.

[20] C. Olaru and L. Wehenkel, A complete fuzzy decision tree technique, *Fuzzy Sets and Systems*, vol.138, no.2, pp.221-254, 2003.

[21] Y. Song and Y. Lu, Decision tree methods: Applications for classification and prediction, *Shanghai Archives of Psychiatry*, vol.27, no.2, pp.130-135, 2015.

[22] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition, Morgan Kaufmann Publishers, 2005.

[23] A. W. Moore, *Naïve Bayes' Classifier*, School of Computer Science, Carnegie Mellon University, 2004.

[24] Y. Oyang, S. Hwang, Y. Ou, C. Chen and Z. Chen, Data classification with radial basis function networks based on a novel kernel density estimation algorithm, *IEEE Trans. Neural Networks*, vol.16, no.1, pp.225-236, 2005.

[25] G. Mountrakis, J. Im and C. Ogole, Support vector machines in remote sensing: A review, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.66, no.3, pp.247-259, 2011.

[26] A. V. Zakharov, M. L. Peach, M. Sitzmann and M. C. Nicklaus, A new approach to radial basis function approximation and its application to QSAR, *Journal of Chemical Information and Modeling*, vol.54, no.3, pp.713-719, 2014.

[27] M. Tsypin and H. Röder, On the reliability of kNN classification, *Proc. of the World Congress on Engineering and Computer Science*, San Francisco, USA, 2007.