Machine Learning Methods for Abnormality Detection in Hard Disk Drive Assembly Process: Bi-LSTM, Wavelet-CNN and SVM

Masayuti Simongyi
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
Masayuti.S@student.chula.ac.th

Prabhas Chongstitvatana
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
Prabhas.C@chula.ac.th

***Abstract*—The research proposes methods to detect abnormality in assembly of a hard disk drive. Three machine learning techniques are employed to classify the drives from assembly process into good and bad class. A good class represent a drive that all components are properly installed while a bad class represent an abnormal drive that some components are missing or improperly installed. The voice coil motor current motor is measured and collected from physical drives in assembly line for using as training and testing data set. Since the amount of bad drives in hard disk drive assembly process are much smaller than the good drive which introduce the imbalance problem during training process, this paper also set the experiment of varying amount of training data set that can satisfy the training in practical hard disk drive assembly process. Bidirectional Long Short-Term Memory, Wavelet transform with Convolutional Neural Network and Support Vector Machine are chosen as proposed machine learning models to classify this task. The comparison between each technique is discussed.**

Keywords— Hard Disk Drive, Voice Coil Motor, Detection, Support Vector Machine, Convolutional Neural Network, LSTM, Bi-LSTM, GoogLeNet.

I. Introduction

In the process of manufacturing and assembling the hard disk drive. It involves the assembly of various hard disk drive components together. Hard disk drive consists of the following main components: voice coil motor (VCM), pivot, ramp and actuator arm, etc. In addition to the main components that make the hard disk drive work properly. The hard disk drive also contains components used to prevent damage to the read/write heads and the disks. For example, a latch has the function of preventing the read/write head from get off garage when the shock occurs. A breather filter acts to prevent leaks of helium gas while writing a servo signal. The breather filter is required to be installed and align with top cover properly otherwise it may be hit by a sliding actuator arm, which is frequently moved in and out, because of the breather filter is mounted on top cover nearby position of read/write head. The collision between the actuator arm and the breather filter not only cause the damage to the head but also produces the particle that circulate inside the hard disk drive. This particle has been proven to be a major cause of head and disk media collision while writing servo signal.

From above, it was found that the problems encountered in assembling the hard disk can be grouped into two categories. The first is that the parts are not completely assembled. Most of them are small pieces that are unnoticeable. For example, a latch which event though it has visual mechanic inspection with human eyes before closing hard disk drive top cover but in some case there still have a problematic hard disk drive that can pass through this check and continue to the next process. The second category of problem is that all parts are completely assembled but some parts are not properly installed, for example, installing a breather filter that mentioned previously. In this study, we label the drives that fall into these categories as a “Bad” class drive and label the remaining drives that pass server writing process as “Good” class, we excluded the drives that are completely and properly installed but fail during servo writing process from this study.

Three Machine learning techniques are employed for detecting this fault. The techniques are applied to the process of writing the servo signals which is done after the hard disk components are completely assembled from the clean room.

The first technique is Bidirectional Long Short-Term Memory (Bi-LSTM), this technique is well-suited to classifying, processing and making decisions based on time series data because it can learn long-term dependency between each step input of sequence data and as VCM current data set is the time series of current during load operation so this technique is promising.

The second technique is Wavelet transform with convolutional neural network using the transfer learning from GoogLeNet. In this study it is difficult to find a sufficient amount of drives to use as training data for a deep CNN since it require a large amount of training data set and it is computational expensive when training it from scratch. So we leverage the existing neural network GoogLeNet that have been train on large data set adapt to our classification technique as pretrained network for image recognition. GoogLeNet is a deep CNNs originally design to classify image in 1000 categories .We reuse the network architecture of CNN to classify binary class of VCM current data based on the images from continuous wavelet transform of VCM current data. GoogLeNe is winner of imageNet Large Scale Visual Recognition Competition (ILSVRC) in 2014 so it is promising to use this pretrained network to our CNN classification.

The last technique that we proposed is Support Vector Machine (SVM) which is well-known for binary classification that give the largest margin between the classes. Usually SVM have a good classification on linearly separable data set and can use a kernel trick to solve that problem that not linearly separable.

978-1-7281-1929-8/18/$31.00 ©2018 IEEE
DOI 10.1109/EECS.2018.00079

Due to the difficulty of finding abnormality drives from assembly process because the failure rate is very small. So this study use much less number of bad drives compare to good drive for training the models. This introduce the imbalance problem to the training process. To tackle with this imbalance problem we create three experiment groups which each group have varying number of training dataset and varying proportion of good and bad drives and we evaluate in in the precision and recall of both classes.

II. Related Works

There are many works that use machine learning techniques to identify defects, faults and abnormalities in systems. Support Vector Machine is one favorited techniques that have been chosen. The algorithm was developed by C. Cortes and V. Vapnik at AT&T Bell Labs [1] to tackled two-group classification problems by computing the best hyperplane that separate all data points of one class from another class. A. Kumar and R. Kumar propose a method to detect defect from vibration signal of centrifugal pump using time-frequency analysis using support vector machine [2]. M. Ebrahimi, M. Khoshtaghaza, S. Minaei, and B. Jamshidi uses SVM technique to detect pest in strawberry greenhouses [3]. Joseph F. Murray, Gordon F. Hughes and Kenneth Kreutz-Delgado [4] use machine learning method for predicting failure in hard disk drive using Self-Monitoring and Reporting Technology (SMART) data. Many techniques have been use in this study including Support Vector Machines and Multiple Instance Naïve Bayes (mi-NB) and they found that SVM with redial base function kernel gave high detection rate with low false alarm when compare to other methods.

Long Short-Term Memory was proposed by S. Hochreiter and J. Schmidhuber [5] which was designed to avoid long-term dependency problem and solve the issue of gradient explosion or diminish in learning process of RNN. LSTM is among the best in learning sequence and time series data and this technique have been applied in various application. J. Li and Y. Shen [6] use Bi-LSTM to generated images description. A. M. Ertugrul and P. Karagoz [7] used Bi-LSTM to classified movie gene from plot summaries. A. H. Mirza and S. Cosan use sequence LSTM to detect computer network intrusion. Cheng Feng et al [8] use LSTM network and package signature to implement multi-level anomaly detection in Industrial Control System (ICS). Y Wang et all [9] use LSTM to predict the water quality. Y. Heryadi et al [10] experiment with various techniques including Stacked LSTM to recognize transaction amount for fraudulent transaction.

A deep convolution neural network is the widely used for image recognition. Y. Le Cun et all [11] proposed to use the technique for handwritten digit recognition. C. Szegedy, W Liu et [12] proposed a deep convolutional neural network architecture code name Inception to classify and detect image in Imagenet Large-Scale Visual Recognition challenge 2014 (ILSVRC14) and won the prized. I. Haberal and H. Ogul [13] use a deep CNN network to make a prediction of protein metal binding-site.

Wavelet transform can be used as feature extraction technique in data. Q. Zhao and L. Zhang [14] use wavelet transform to extract features from The electrocardiogram (ECG) signals. R. F. Leonarduzzi [15]. T. Li and M.Zhou [16] classified EGC signals using Wavelet Packet Entropy and Random Forests.

III. Hard Disk Drive Assembly Process

The hard disk drive assembly process must be done in a clean room. After the components are cleaned, the first step is to install the disk media and then the disk clamp to hold the disk media together. After that, it will be verified that if the hard disk drive media has been properly installed before installing the ram and other devices, as shown in Fig. 1. It can be seen that in the hard disk drive assembly process it is not only involve the installation of each component together but it is also include of the verification of the component that has been installed. For example, the verification of disk media balance and the leak test but there are also a number of components that could not be verified. For example, a ram, voice coil motor, actuator arm, read/write head and a latch. This is because there is no proper way to verify these components better than visual mechanical inspection.

After all the hard disk drive components are assembled, they are closed with the top cover and passed to the next process with printed circuit board assembly (PCBA). It's ready for supply the voltage and the reader is pushed into the disk media to write a servo signal.

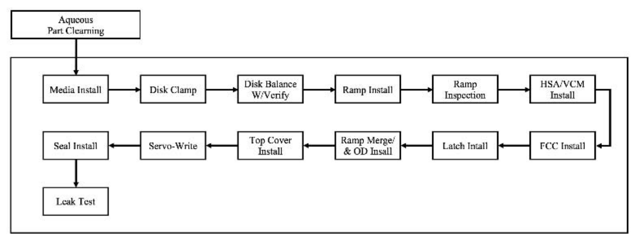


Fig. 1. Process of assembly hard disk drive in clean room.

In order to read and write servo signal on disk media, the first step is to spin up spindle motor to make it spin up at constant velocity and then loading the heads onto the disk. If the head is loaded while the spindle motor is not spinning at our target speed it has the potential to cause the read/write head crash with the disk media. Generally, in the firmware of the hard disk drive, it is always necessary to have this verification step before loading head onto the disk.

When the spindle motor is spinning at speed, the current is supplied to voice coil motor so that the actuator arm is pushed onto the disk. This process requires a different level of current because the force that need to push actuator out of garage until get actuator onto the disk is varying and because of the voice coil motor itself is a like coil which generate the back EMF when supply the current through it. So, it is important to have close loop system built in as shown in Fig. 2 to accurately control the speed of the actuator.

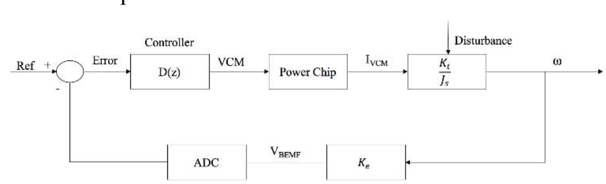
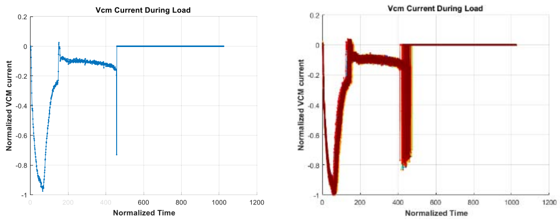


Fig. 2. Closed-loop control system of actuator arm.

The current that supply to voice coil motor since head parking on the garage until head successfully load onto disk media is plot and shown in Fig. 3(A). At the beginning, enough amount of current need to be supplied to make actuator out of garage and latch and then the current needed is gradually reduce until close to zero when head is on disk media and detect the crash stop.

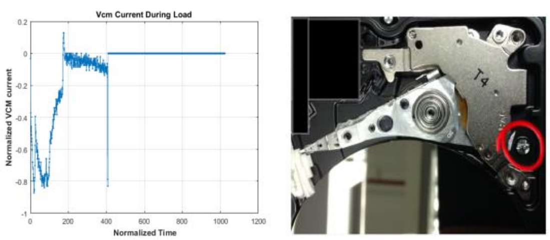
When we collect voice coil motor current during loading head operation over 1000 times, the voice coil motor current has a consistent shape as shown in Fig.3 (B) this is because the current that supply to the coil has feedback control as described above.



(A) (B)

Fig. 3. (A) Single VCM current. (B) 1000x overlay of VCM current.

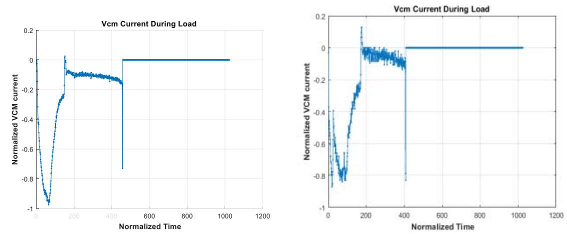
The current that supply to the voice coil motor is continuous value but the current that collect for experiment is sampling from continuous value. This sampling time is equal to the interrupt time for close loop feedback control system. In this study, we allocate memory of 1024 values for each load operation. This will include interval time of 1024 time unit which enough for our validation.



(A) (B)

Fig. 4. (A) VCM current of missing latch drive. (B) Missing Latch position.

From the experiment with the problematic drives it is found that the form of voice coil motor current is change when experiment with incomplete assembly drive. For example, when latch is not installed as shown in Fig. 4(A) the shape of the voice coil current is different from Fig.3 (A). Fig. 4(B) show the position where latch I not installed. Figure 5 show the comparison of VCM current between good drive, Fig. 5(A) and bad drive, Fig. 5(B). From the figure, the shape of VCM current show big difference in the time unit from 0 to 200 which is the time when heads get off the garage and load onto media. Later on, the difference become small during heads fly onto the media and find crash stop. This is because the missing components of the drive such as latch influence the VCM current only on early time during heads get off from the garage, it hold actuator to prevent actuator unexpectedly load heads onto the disks. After actuator can get off from garage, latch has no longer influence to the VCM current so the differences are decreased.



(A) (B)

Fig. 5. (A) VCM current of good drive. (B) VCM current of bad drive.

IV. Research Method

A. Abbreviations and Acronyms

In the process of collecting data of voice coil motor current to use for training the models, the data is stored in pair format between voice coil motor current data and the class of assembly process. This class indicates that which hard disk drives are assembled correctly. The Good class represents the hard disk drive that assembly correctly and the Bad class represents the hard disk drive that have abnormality in hard disk drive assembly process. We excluded the drives that are completely and properly installed but fail during servo writing process from this study. The voice coil motor current data is sampled every time unit starting when load command is invoked for 1024 time units. This time interval is long enough to cover event from loading head until head loading onto media and found the crash stop, on average this process takes around 300 to 600 time unit as depicted in Fig. 3(B).

The voice coil motor current data from many drives are collected in form of Fig. 6 in which Fig.6(A) is array that each row collect the data of voice coil motor from 1 time unit to 1024 time unit in Fig. 6(B) in each row collect the class of hard disk drive

	1	2	3	4	.	.	.	.	.	1024
HDD#1 =>										
HDD#2 =>										
HDD#3 =>										

(A)

	Class
HDD#1 =>	Bad
HDD#2 =>	Bad
HDD#3 =>	Good

(B)

Fig. 6. (A) array of VCM current. (B) class of hard disk drive.

B. Experiment Groups

In this study, the voice coil motor current data have been collected from 7539 drives. In which only 20 abnormal drives (Bad class) is collected , the remaining data set, 7519 is collected from good drive (Good Class). The small ratio of bad class to good class samples (0.27%) introduce imbalance problem during training process. So we divide the collected data set into three experiment groups and in each group is separate for testing and validation data set. Table 1 shows the data set for each experiment group. For each experiment group it has different number of training data set, validation data set and ration of bad drive in training data set. The ratio of bad drive is calculated by the number of bad sample in training data set divide by the total number of training data set. The experiment also plays with varying number of training data set because in practical it is needed to consider how many training data set should be collected from assembly line to adequately satisfy the training requirement of each machine learning model. Larger data set need more time and resources for collecting while fewer data set may introduce overfitting to the machine learning model. The number training data set for experiment group #1, #2 and #2 is 500, 100 and 50 respectively and the ratio of bad drive in each experiment group is 3%, 10% and 10% respectively.

TABLE I : DATA SET SEPARATED BY GROUP OF EXPERIMENT.

Experiment Group #	Testing Data Set		Validation Data Set		Total	Ratio of Bad Drive
	Good	Bad	Good	Bad		
#1	485	15	7034	5	500	3%
#2	90	10	7429	10	100	10%
#3	45	5	7474	15	50	10%

C. Training and Validation

All the techniques that we have proposed in section V which are Bi-LSTM, a deep CNN with transfer learning from GoogLeNet and SVM are supervised learning technique which is the learning technique that map input to an output based on example of input and output pairs. So it is required to separate VCM current data set into two data set, one is training data set which use to train the network model and second is validation data set. The network model that that we get from training data set will be use as model to evaluate validation data set. The classified result from validation data set with the models are compare with real HDD assembly class. The performance metrics are evaluated. The process flow is shown in Fig. 7.

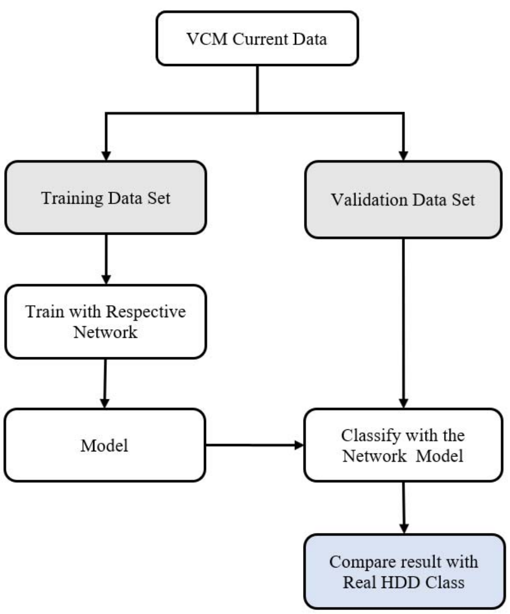


Fig. 7. Testing and validation work flow.

D. Performance Measurement

The Precision is defined as in Eqs. (1), where TP (true positives) referred to the number of correctly predicted positive sample (bad drive). FP (false positive) is referred to the number of negative samples (good drives) that are incorrectly predicted as positive sample (bad drive). The Recall is defined as Eqs. (2), where FN (false negative) is the number of positive samples that are incorrectly predicted as negative. The accuracy is computed as in Eqs. (3).

In this study, due to the number of negative sample is far more than the number of positive sample which introduce imbalance issue to training process, so the F-measure is evaluated. The F-measure is the harmonic average of precision and recall which be written as Eqs. (4).

$$\text{Precision} = TP/(TP+FP) \quad (1)$$

$$\text{Rcall} = TP/(TP+FN) \quad (2)$$

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (3)$$

$$\text{F-Measure} = 2TP/(2TP+FP+FN) \quad (4)$$

V. PROPOSE METHODS

A. Bidirectional LSTM (Bi-LSTM)

Long Short Term Memory (LSTM) network is a special kind of Recurrent Neural Network (RNN) which can learn long-term dependency between each step input of sequence data and it is designed to avoid long-term dependency problem and solve the issue of gradient explosion or diminish due to long time lags during backpropagated error in learning process of RNN. So LSTM is well-suited to classify and predict long-time series data.

In this study, we use Bidirectional LSTM (Bi-LSTM) which is the variation of LSTM network that can learn both forward information and backward information of VCM current data at any particular time. The VCM current input is feed into Bi-LSTM network as show in Fig. 8 , where h_t and c_t represents the output and cell state of Bi-LSTM cell at time t respectively, and follow by Bi-LSTM layer which have hidden layer that compound of 100 hidden units. To predict the class label, the network end with fully connected layer, a softmax layer and classification output layer as depicted in Fig. 9.

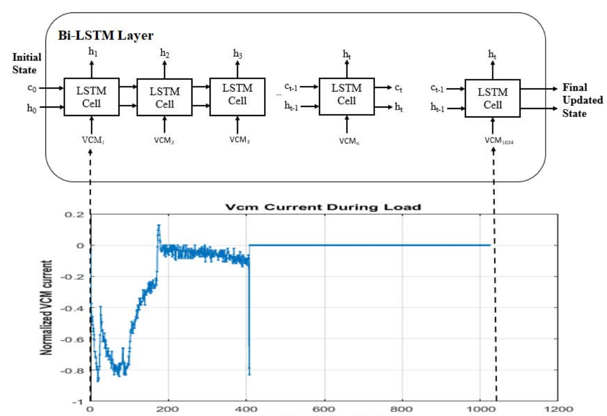


Fig. 8. LSTM Network Layer

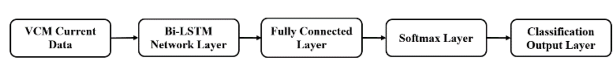


Fig. 9. LSTM Network Architecture

In this study, The Bi-LSTM network use stochastic gradient descent with momentum during training with the momentum value of 0.9 and initial learning rate of 0.01.

B. Wavelet Transfrom with Convolutional Neural Network

In this study we apply a deep convolutional neural network (CNN) GoogLeNet, a pretrained network for image recognition to classify VCM current data of good and bad drives base on a time-frequency representation using the continuous wavelet transform (CWT), we shortly call this method as Wavelet-CNN.

GoogLeNet is a deep convolutional neural network designed for classify images of 1,000 categories. In this study we leverage this CNN to classify the abnormality of the hard disk drive assembly process based on images from CWT of VCM current data.

CWT is a time-frequency transformation that is widely use in image compression because it is provides significant improvements in picture quality at higher compression ratios over conventional techniques. It is also use in acoustics processing and pattern recognition because it good to detect abrupt changes in the signal.

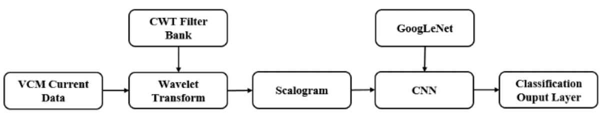
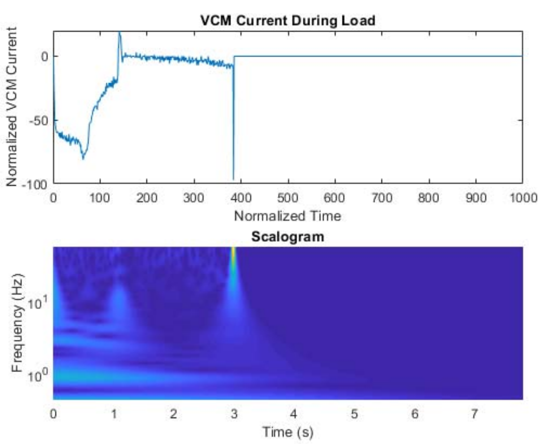
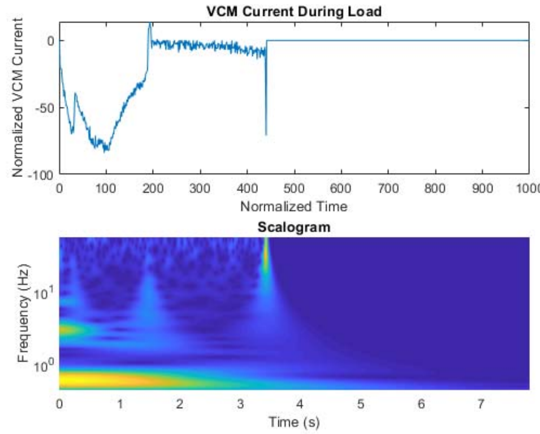


Fig. 10. Wavelet-CNN Network Architecture

To create time-frequency representations of VCM current, we need to transform the data into representations call "scalogram". A scalogram is the absolute value of the CWT coefficients of the data which can be precompute by a CWT filter bank. After create CWT filter bank with 1024 sample. We use the filter bank to take the CWT of the first 1024 samples of VCM current data and obtain the scalogram from the coefficients as shown in Fig. 10. The comparison of scalogram between good and bad drive is shown in Fig. 11.



(A)



(B)

Fig. 11. Scalogram of a good drive (A) and a bad drive (B).

To be compatible with the GoogLeNet architecture the images that to be converted to RGB of size 224-by-224-by-3.

In GoogLeNet Network layer, the earlier layers handle more common feature of images such as edge and color. Since these layers are common for all images so we don't need to changes it. However, we need to modified the later layer of the network which focus more on specific feature to well-suited our needs.

We modified the dropout layer 'pool5-drop_7x7_s1' which randomly set input elements to zero with a given probability to help prevent overfitting of training data. The default probability value of this layer is 0.5 we change it to 0.6. We set the fully connected layer from default value of 1000 (1000 categories of images) to 2 which equal to our classification lass (Good and Bad Class).

During training, stochastic gradient descent algorithm is used. In each iteration the gradient of loss function is evaluated and the weights are updated. We set our initial learning rate to 0.0001 and momentum to 0.09.

Since the number of training samples are vary for each experiment groups. The first experiment group which handle only 50 training sample can be train individually. However, for the larger training sample we need to group some sample as mini batch to be a subset of the training set to use in each iteration. We choose the number of maximum epoch to100as the loss for this value is close to zero, to prevent underfitting model. Increasing the maximum epoch more than 100 may cause the problem of overfitting. One epoch is a full pass of the training algorithm over the entire training set.

C. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a pattern classification algorithm which can be used for classification and regression. SVM classifieds data by calculating the optimal hyperplanes that separate all data points of one class from another class. The optimal hyperplane for SVM is the one that have largest margin between the two classes, which is the maximum width of the parallel line in the hyperplane that has no interior data points as depicted in Fig. 12 which show the classification of Good and Bad class using SVM.

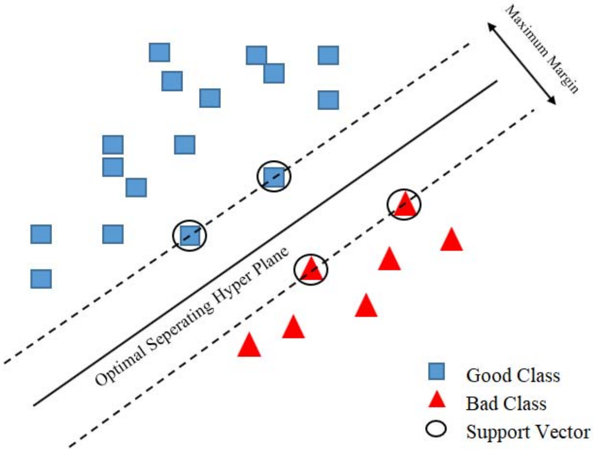


Fig. 12. Support Vector Machine for classification

If the data is linearly separable then the optimal separating hyper plane is existing and can be described as in Eqs. (5) where N is the number of training data of data point (x_i, y_i), x_i is the input data and $y_i \in \{+1, -1\}$ corresponds to its target value of the output class, i.e. Good class equal to +1 and Bad class is equal = -1. The parameter w and b is the weight and bias of the SVM respectively.

$$f(x) = w^T x + b = \sum_{i=1}^{N} w_i x_i + b = 0 \qquad (5)$$

In case that the data is non-linear in lower dimensional space it can be transform to higher dimensional space using kernel trick, which is incorporate kernel function for computing dot product of mapping function. The kernel function is $K(x_i, x) = x_i^T x$ for Linear Kernel $K(x_i, x) = (x_i^T x + 1)^d$ for Polynomial Kernel and $K(x_i, x) = e^{(-\gamma||x - x_i||^2)}$ for Radial basis function Kernel (RBF).

In this study, we have experimented with many kernels and found that Linear Kernel give the best result for this classification. So we use SVM with Linear Kernel as our technique to compare with other machine techniques that we proposed in this paper.

VI. Experiment Result and Discussion

This study was evaluated with validation data set as described in section 4. An Accuracy, precision, recall and F-measure were calculated as performance measurement.

The results were evaluated at the end of training epoch. In this study we set the maximum epoch to 100 for Bi-LSTM and Wavelet-CNN network model. However, we observed that training accuracy and training loss were converged earlier than that, around 30-40 epochs. The validation results are shown in Table II and its performance measurement are shown in Table III.

For experiment group #1, which have 7034 negative samples (Good drive) and 5 positive samples (Bad drive). Bi-LSTM based prediction and SVM based prediction show perfect results in matric measurement which resulted in 100% accuracy, 100% precision, 100% recall and 100% F-Measure while Wavelet-CNN based prediction resulted in 99.94% accuracy, 55.6% precision, 100% recall and 71.4% F-measure.

For experiment group#2, which have 7429 negative sample and 10 positive samples. Bi-LSTM based prediction still shown perfect matric measurement performance, 100% accuracy, 100% precision, 100% recall and 100% F-measure while SVM based prediction show a lightly dropped in accuracy, dramatically dropped in precision and F-measure which resulted in 99.93% accuracy, 66.7% precision, 100 % recall and 80% F-measure. Wavelet-CNN shown worsen result than its experiment's result in group#1 which resulted in 99.80% accuracy, 40% precision, 100% recall and 57.1% F-measure.

For experiment group#3, which use totally 50 samples as training data set as described in section 4B. In this experiment group, 7474 negative samples and 15 positive sample were evaluated. None model can archive perfect matric measurement performance. Bi-LSTM based prediction resulted in 99.97% accuracy, 100% precision, 86.7% recall and 93.0% F-measure. Wavelet-CNN resulted in 99.97% accuracy, 46.4% precision, 86.7% recall and 60.5% F-measure. SVM based prediction resulted in 99.96% accuracy, 46.4% precision, 86.7% recall and 60.4% F-measure.

The results for each proposed method are compared across all experiments groups. Fig. 13 show the accuracy evaluation matric compare the across the proposed machine learning methods and experiment groups. The comparison of precision, recall and F-measure are shown in Fig. 14, 15 and 16 respectively.

TABLE II : Classification result from validation data set

ML Techniques	Groups	Good	Bad	TP	FP	TN	FN
Bi-LSTM	#1	7034	5	5	0	7034	0
	#2	7429	10	10	0	7424	0
	#3	7474	15	13	0	7474	2
Wavelet-CNN	#1	7034	5	5	4	7030	0
	#2	7429	10	10	15	7414	0
	#3	7474	15	13	15	7459	2
SVM	#1	7034	5	5	0	7034	0
	#2	7429	10	10	5	7424	0
	#3	7474	15	15	3	7471	0

TABLE III : PERFORMANCE MEASURMENT MATRICS FSAFS

ML Techniques	Groups	Accuracy	Precision	Recall	F-measure
Bi-LSTM	#1	100.00%	100%	100%	100%
	#2	100.00%	100%	100%	100%
	#3	99.97%	100%	87.0%	93.0%
Wavelet-CNN	#1	99.94%	55.6%	100%	71.4%
	#2	99.80%	40.0%	100%	57.1%
	#3	99.77%	46.4%	86.7%	60.5%
SVM	#1	100.00%	100%	100%	100%
	#2	99.93%	66.7%	100%	80.0%
	#3	99.96%	83.3%	100%	90.9%

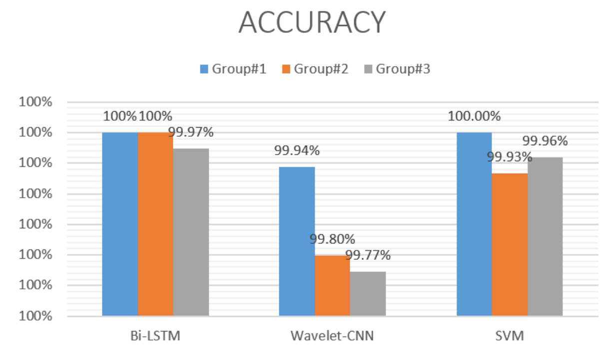


Fig. 13. Accuracy evaluation matric using validation data set.

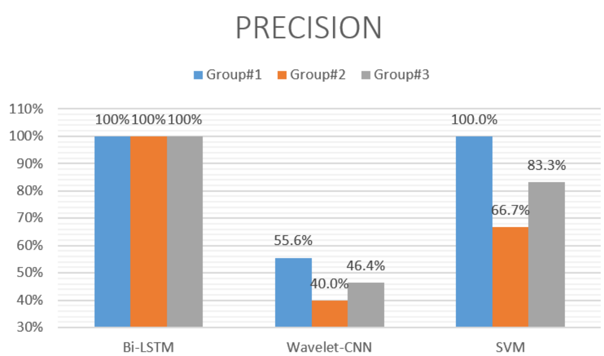


Fig. 14. Precision evaluation matric using validation data set.

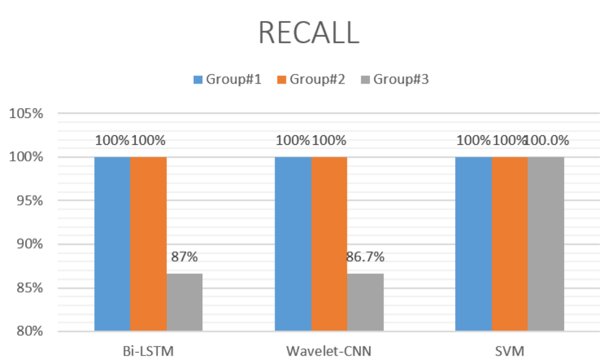


Fig. 15. Recall evaluation matric using validation data set.

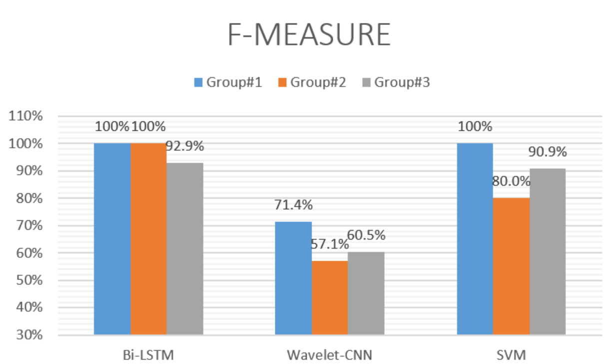


Fig. 16. F-Measure evaluation matric using validation data set.

VII. Conclusion

We presented three methods to detect abnormality in hard disk drive assembly process which is Bi-LSTM, wavelet-CNN and SVM. From the experiments results, we found that Bi-LSTM and SVM are a powerful approach for detect this abnormality since it's give perfect performance measurement matric (100% accuracy, 100% precision, 100% recall and 100% F-measure) on experiment group #1 and #2.

Even though, there is no any methods that can archive perfect performance measurements on experiment group # 3, Bi-LSTM show highest accuracy (99.97%) over SVM (99.96%) and Wavelet-CNN (99.77%) and Bi-LSTM still resulted perfect score on precision in experiment group #2 while SVM and Wavelet-CNN resulted only 83.3% and 46.4% respectively. However, SVM show better recall result than Bi-LSTM on this experiment group, 100% over 87%.

In practical application of this detection in hard disk drive manufacturing process. If the number of training samples are difficult to collected as in experiment group # 3, one need to trade-off between two consequences. One is overkilled the good drive (fail FP drives) and second is letting the bad drives pass through the customer (not fail FN drive).

From the experiment group #3, the FP rate that classify by SVM is 3/7489 which is around 400 drives in a million of manufacturing drives. This mean that 400 drives would be over-killed if we employ SVM in experiment group #3 (This value is 0 when classify with Bi-LSTM). On the other hand, if we use Bi-LSTM model to classify, the recall with this model is 86.7% which mean if there is 100 of abnormality drives in a million of manufacturing drives. There will be 14 drives pass the abnormality detection (This value is 0 when classify with Bi-SVM).

From our discussion above, we encourage to use training sample size at least 100 samples with 10% of positive drive in it to perfectly classify the good and bad drive with Bi-LSTM and SVM model.

The advantage that SVM have over Bi-LSTM is the time to train the model. It took around 3 hours to train Bi-LSTM model with 500 samples size, but it took less than 5 minutes to train the same sample with SVM model.

The result from Wavelet-CNN show worsen than other model in performance measurement matrices, this maybe because we not directly training the model from scratch, but we leverage the existing GoogLeNet as our pretrained model for CNN. GoogLeNet is good in classify general images like animals, flower etc. because it is trained based on these images but not well-suited to classify the images in our study.

References

[1] C. Cortes and V. Vapnik. "Support vector networks. Machine Learning," 2O:l-25, 1995.

[2] A. Kumar and R. Kumar, "Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump," Measurement, vol. 108, no. Supplement C, pp. 119 – 133, 2017.

[3] M. Ebrahimi, M. Khoshtaghaza, S. Minaei, and B. Jamshidi, "Vision-based pest detection based on svm classification method," Computers and Electronics in Agriculture, vol. 137, no. Supplement C, pp. 52 – 58, 2017.

[4] Joseph F. Murray, Gordon F. Hughes, Kenneth Kreutz-Delgado, "Machine Learning Methods for Predicting Failures in Hard Drives: A

Multiple-Instance Application" Journal of Machine Learning Research, vol. 6, pp 783–816, 2005.

[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory" Neural Compute, vol. 9, no. 8, pp. 1735-1780, 1997.

[6] J. Li and Y. Shen, "Image describing based on bidirectional LSTM and improved sequence sampling," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, 2017, pp. 735-739.

[7] A. M. Ertugrul and P. Karagoz, "Movie Genre Classification from Plot Summaries Using Bidirectional LSTM," 2018 IEEE 12th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, 2018, pp. 248-251.

[8] C. Feng, T. Li and D. Chana, "Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks," 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Denver, CO, 2017, pp. 261-272.

[9] Y. Wang, J. Zhou, K. Chen, Y. Wang and L. Liu, "Water quality prediction method based on LSTM neural network," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017, pp. 1-5.

[10] Y. Heryadi and H. L. H. S. Warnars, "Learning temporal representation of transaction amount for fraudulent transaction recognition using CNN, Stacked LSTM, and CNN-LSTM," 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Phuket, 2017, pp. 84-89.

[11] Y. Le Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson, "Handwritten digit recognition with a back-propagation network," Advances in Neural Information Processing Systems 2, San Francisco, CA, USA, 1990, pp. 396-404.

[12] Szegedy, Christian et al. "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015), pp. 1-9.

[13] İ. Haberal and H. Oğul, "DeepMBS: Prediction of Protein Metal Binding-Site Using Deep Learning Networks," 2017 Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), Corfu, 2017, pp. 21-25.

[14] Zhao, Q. and Zhang, L.,. ECG feature extraction and classification using wavelet transform and support vector machines. IEEE International Conference on Neural Networks and Brain,2, 2005, pp. 1089-1092.

[15] R. F. Leonarduzzi, G. Schlotthauer and M. E. Torres, "Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, 2010, pp. 110-113.

[16] Li, T. and Zhou, M., "ECG classification using wavelet packet entropy and random forests," Entropy, 18(8), 2016, p.285.