Implementation of Unsupervised Machine Learning Techniques for Monitoring Coal Mills at a South African Power Plant

Sekhoto, Dorothea Lawretta Makomane

(出版者 / Publisher)
法政大学大学院理工学研究科

(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学研究科編

(巻 / Volume)
65

(開始ページ / Start Page)
1

(終了ページ / End Page)
7

(発行年 / Year)
2024-03-24

(URL)
https://doi.org/10.15002/00030725

Implementation of Unsupervised Machine Learning Techniques for Monitoring Coal Mills at a South African Power Plant

Name: *Makomane Dorothea Lawretta Sekhoto*

Major: Applied Informatics, Graduate School of Science and Engineering,

Hosei University,

Supervisor: Dr. Akihiro Fujii

***Abstract*— This study uses advanced machine learning techniques to create and assess effective predictive maintenance strategies for coal mills. The focus is determining coal mills' safe and unsafe operational phases, predicting potential tripping events, and scheduling maintenance to prevent equipment failures. The study used Gaussian Mixture Models (GMM) to cluster operational phases and Long Short-Term Memory (LSTM) models, including the CNN+LSTM architecture, for predictive analysis. The study evaluated the models using various metrics such as accuracy, precision, recall, F1, and ROC_AUC score. The research findings show that LSTM models, especially the CNN+LSTM architecture, can accurately identify unsafe operational phases and predict coal mill tripping events. The CNN+LSTM architecture has an output value that increases towards the failure moment, indicating its effectiveness in capturing relevant temporal patterns. This can be useful for scheduling inspections before failure.**

Keywords: Coal Mills, Clustering, Gaussian Mixture Model (GMM), Unsupervised Machine Learning, Long Short-Term Memory (LSTM)

I. INTRODUCTION

In South Africa, coal is the primary source of electricity generation. The country has several power plants that account for most of its electricity production. Coal-fired power plants make up 75% of South Africa's electricity generation. Coal mills are a crucial component in the operation of coal-fired power plants, as they are responsible for grinding coal into small particles for the combustion process and electricity generation. Coal mills (Fig. 1) need to always operate efficiently; to ensure optimal operation, research has focused on analyzing the effects of improper mill load line air-fuel ratio and other factors that can impact coal mill performance [1]. However, limited methodologies are available for classifying or labeling unsupervised coal mill operations data. This study aims to contribute to a better understanding of large and complex data of coal mill systems.

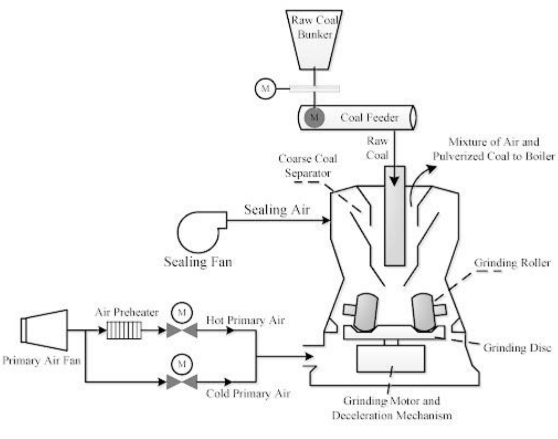

Fig. 1: The vertical spindle MPS spindle Coal mill operation [1]

Coal-fired power plants rely on coal mills to crush, dry, and transport coal to the boiler burners, where it is burned in a combustion process to generate electricity [2]. These mills are crucial for ensuring the correct particle size of the coal and maintaining plant efficiency. However, complex issues may negatively impact mill performance, which engineers and operators must address to optimize plant operation. It is crucial to ensure that coal mills operate safely and efficiently to avoid power loss or complete

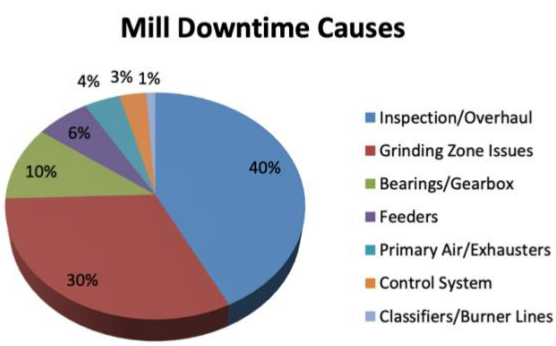

Fig. 2: Mill downtime causes [4]

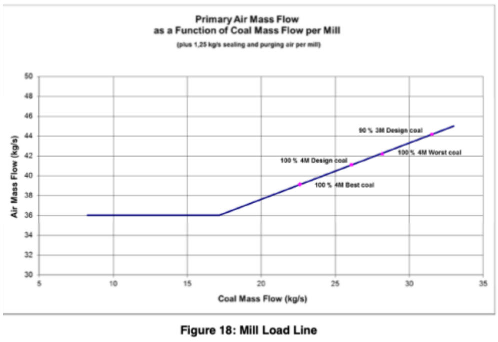

Fig. 3: The Primary Air Fuel Design Diagram [5]

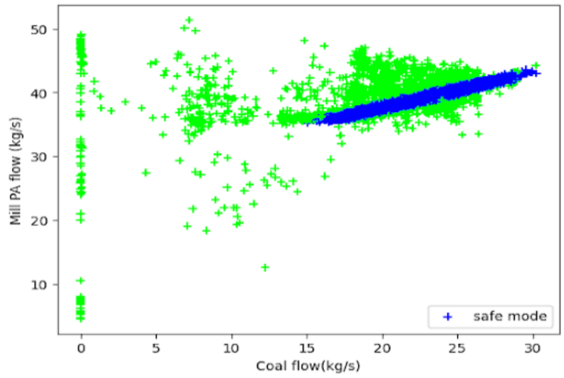

Fig. 4: Coal flow(kg/s) vs. Mill Primary Air flow(kg/s)

shutdown (Fig. 2) of the plant unit, which can increase harmful emissions [3]. Optimizing coal mill performance and reducing maintenance downtime is essential, and several factors (Fig. 2) can affect mill performance, such as grinding zone issues, bearing/gearboxes, feeders, primary air or exhausters, control systems, and classifiers issues [4].

To effectively handle mechanical failures in engineering systems, it's crucial to have a carefully thought-out maintenance plan (Fig. 5). By selecting the right maintenance approach based on the type of equipment or system, it's possible to identify losses early on and take prompt action, which leads to optimal resource utilization and better planning [10].

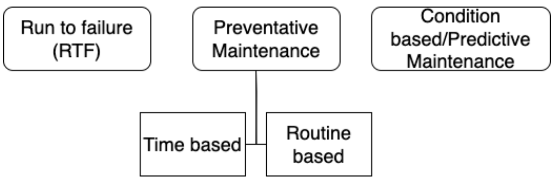

Fig. 5: Traditional maintenance strategies

The dataset used in the research is from a medium-speed coal mill used in an 800MW power station in South Africa. The research will focus on coal flow, primary airflow, mill current, and inlet pressure. These variables are crucial for feeding and drying the coal in the mill:

- **Coal flow (kg/s):** The amount of coal fed into the mill depends on the boiler load demand.
- **Primary Air (PA) flow (kg/s)**: The PA flow is directly related to the coal load demand and is responsible for conveying and drying the coal as it is fed through the mill.
- **Mill current (A)**: The current motor that drives the rotating grinding track.
- **Mill inlet pressure (kPa)**: The inlet pressure is formed by a combination of the PA flow and the settings of the cold/hot air dampers, and it is responsible for feeding and drying the coal.

To ensure that mills operate efficiently between 40% to 100% boiler load, it's important to consider variables that affect coal mill performance. The coal flow into the mill determines the primary air mass flow, which is controlled by the characteristic curve in Fig. 3. The measured PA flow is compared to the target value, and any difference is adjusted through the hot/cold air control damper. Getting the air-fuel mixture right is crucial for optimal efficiency. Coal fired Power Plants' smooth running depend mainly on the effective operation of coal mills.

Various theories have been proposed for detecting and diagnosing faults in coal mills [6], including genetic algorithms[7], data-based [8] and model-based methods [9], and deep learning techniques such as Stacked Autoencoders [11] and Long-Short-Term-Memory Auto-Encoder [12]. Other approaches not specific to coal mills but are related to coal mills include Kernel Principal Components Analysis [13], Hierarchical Clustering [14], and Support Vector Regression [14]. The research on classifying operations based on significant input variables was inspired by *W Fan et al*. [15] 's use of a combination of the Gaussian Mixture Model (GMM) and Multi-output Relevance Vector Regression (MRVR) to detect faults and identify operating modes in a coal mill. This innovative approach has shown promising results and holds great potential for improving the efficiency and safety of industrial processes.

This study aims to detect extreme modes of operation and monitor the deterioration of a coal mill over time. Its objective is to tackle the difficulty of labeling a large amount of data in coal mill operations. Furthermore, the study will compare LSTM architectures to GMM-labeled datasets for detecting faults in unlabeled datasets and consider the temporal dependency.

II. METHODS

A. *Overview of the dataset*

The data from a coal-fired coal mill in South Africa was analyzed for two years. After data preprocessing and feature selection, the remaining dataset was explored to identify patterns and clusters. Based on sensor availability, expert knowledge, and correlation analysis, the original set of fourteen variables was reduced to eight. Strongly correlated variables were selected as input to the GMM model (Table I).

B. Data Cleaning and Pre-processing

To ensure accurate analysis, missing data from non-operational periods were removed. Mean values were used to fill in missing data. Normalization and scaling were crucial in preparing data for input into LSTM models. The correlation analysis showed that Mill PA flow, Mill Inlet pressure, and Motor Current were closely related to Coal flow (Table I).

Table I: Top 4 features Importance and odd ratios

Feature	Unit	Importance	Odds ratio
Coal flow	*kg/s*	0.623844	3.199737
Mill PA flow	*kg/s*	0.328844	0.275644
Mill Inlet Pressure	*kPa*	0.027189	0.507782
Motor Current	*A*	0.020124	1.000744

The data shows a wide range of coal mill operations and interdependent variables affected by changes in load demand. Correlated features were easily identified and grouped together.

C. Gaussian Mixture Model

The GMM clusters are formed by assuming a Gaussian shape of chosen features and have been helpful in monitoring coal mill operating conditions [15]. It can identify extreme modes of operation, which can be either safe or unsafe (Fig. 4). However, identifying the optimal number of clusters can be challenging, especially with high-dimensional data. The proposed GMM with EM method is shown in Fig. 6.

Let $X = [x_1, x_2, x_3 \ldots x_m]^T$

$$\mathrm{N}\left(\mathrm{x}\middle|(\mu_k, \Sigma_k)\right) = \frac{1}{(2\pi_k)^{\frac{m}{2}}|\Sigma_k|^{\frac{1}{2}}} \exp^{\left\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right\}} \quad (1)$$

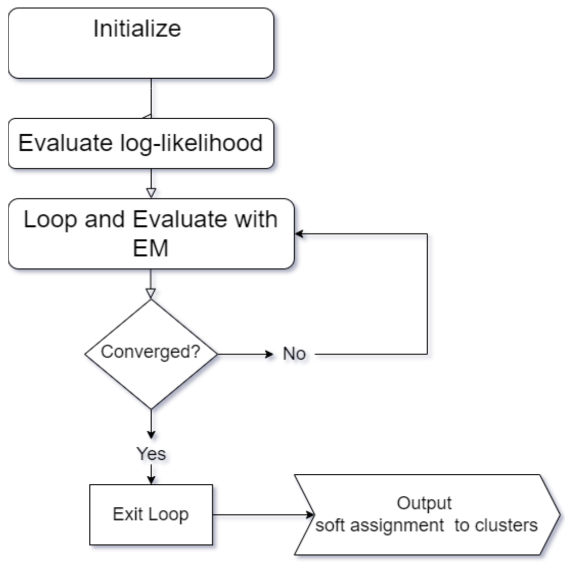

Fig. 6: The proposed GMM with EM method

Where K represents the number of clusters formed by Gaussian Mixtures clusters, covariance (Σ_k), mean (μ_k) and weight (π_k) [16][17]. Where $\theta = \mu_k, \Sigma_k$ the Maximum log probability with weights and covariance.

$$\mathrm{Log}p(X|\theta) = \sum_{k=1}^{k} log \sum_{k=1}^{k} \pi_k N(x_i|\mu_k, \Sigma_k) \quad (2)$$

To use the expectation maximization technique, estimating the number of clusters with K-means initialization. Then refine accuracy with the Expectation-Maximization (Fig. 6) method and solve Equations 1 to 2. Soft assignment is completed with output hyperparameters.

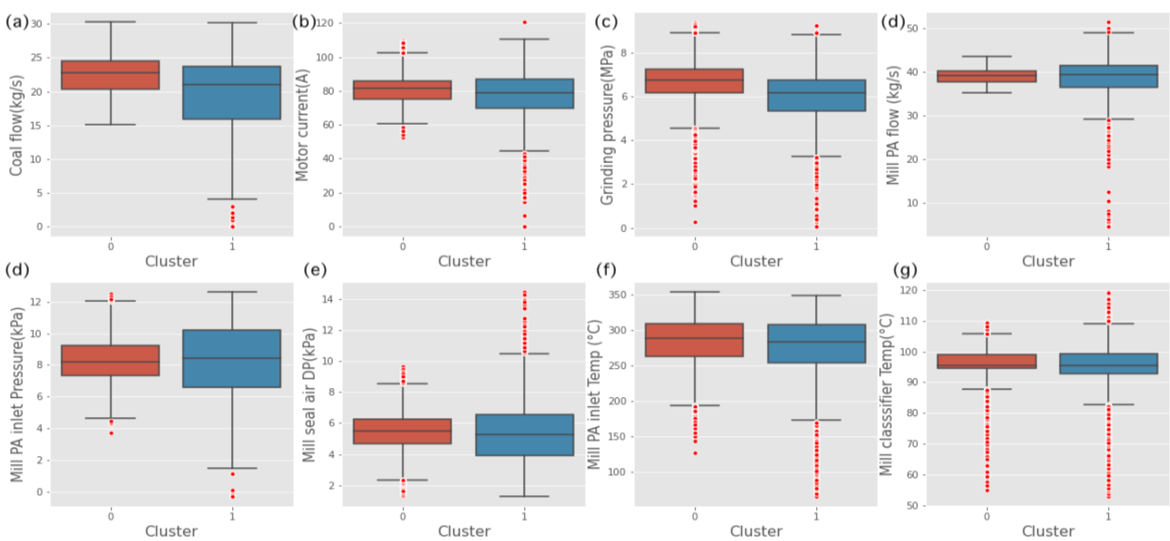

Fig. 7: Box plots of Clusters (0: safe operation and 1): Unsafe operation for feature input variables

D. Cluster Evaluation Criteria

Clustering methods like Silhouette, BIC, AIC, and Davies-Bouldin Index were used to determine the ideal number of hidden clusters. The BIC and AIC are the best methods for selecting the precise number of clusters in GMM [16]. They help in accurately matching the optimal cluster numbers [18].

$$BIC = \log(N)\,\mathrm{X_m} - 2\log(MLL) \quad (3)$$
$$AIC = 2\mathrm{X_m} - 2\log(MLL) \quad (4)$$

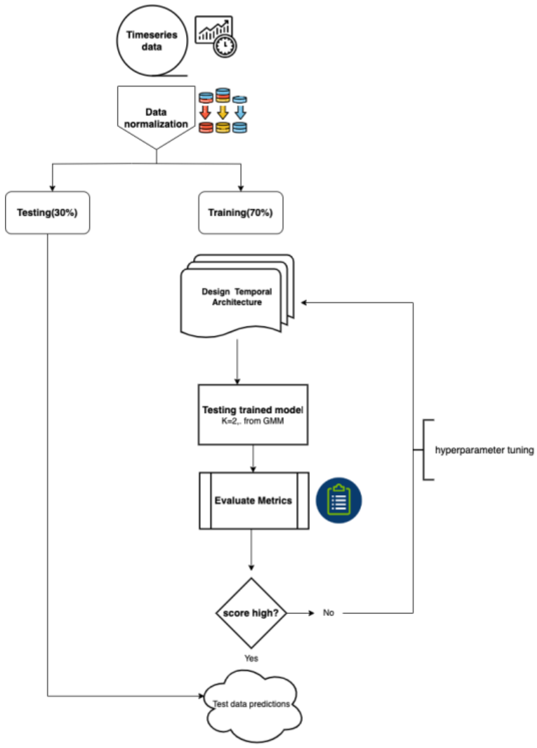

Fig. 8: The process for LSTM architectures

N is the sample size and, MLL is the Maximum log likelihood, X_m is the number of observed features selected (Table I).

To distinguish the clusters, choosing the lower score that shows the similarity between each cluster and its closest one is best [19]. As shown in Equation (7), the Davis-Bouldin score determines the clusters' similarity level [20].

$$\delta_k = \sqrt{\frac{\sum_{n\in\mathbb{R}} \|X_n - C_k\|^2}{N_k}} \quad (5)$$

$$S_{kl} = \frac{\delta_k + \delta_l}{\|C_k - C_l\|} \quad (6)$$

$$DBI = \frac{1}{K}\sum Max \quad S_k \quad (7)$$

Where C_k is the cluster center, X_n is a set of datapoints that belong to cluster K, and $1/N_k$ is the normalization of the number of data points in the cluster.

E. Analyze with design specifications.

To ensure reliability, a safe cluster was identified using GMM (Fig. 4). It was compared to the original design (Fig. 3) to ensure accuracy and optimal performance.

F. The Choice of input X features

K-means was initially used, and other clustering techniques were later employed. However, reducing the number of input features (Table I) had a negative impact on the likelihood score and convergence time. Increasing the number of clusters while keeping the features constant isolated the unsafe mode and subdivided the safe mode into different operational modes. This method is similar to techniques used in previous studies to identify operating modes [17][18]. This study aims to identify unsafe modes of operation and prevent them from occurring.

G. Temporal dependencies models

Multiple LSTM models (Fig. 8) will be presented and evaluated based on their ability to address temporal dependencies using appropriate metrics. The performance of these models will showcase their potential to aid engineers and operators in identifying risks and taking corrective action. A hypothetical scenario will also be discussed to demonstrate how an LSTM model can suggest an inspection to prevent a coal mill failure based on past events.

H. LSTM Evaluation Criteria

After employing GMM clusters, the classification model's performance will be evaluated using the assessment metrics displayed in Table II.

Table II: Evaluation Metrics for classification [21][22]

Accuracy	*P=Precision*	*R= Recall*	*F1 score*
$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$\frac{2\times P\times R}{P+R}$

III. RESULTS

A. GMM Cluster Distribution Analysis

The GMM method can divide a large dataset into safe and unsafe modes. Following the mill load line (Fig. 3), the coal mill can operate safely. Coal flow rate and primary air flow are highly correlated (Fig. 7(a)and (d)), resulting in a tight cluster of safe operating modes. Mill inlet pressure and motor current have a high odds ratio (Table I) of falling into the unsafe operating zone.

B. Cluster Evaluation Criteria

The DBI score is lowest when there are only 2 or 5 clusters (Fig. 9). It's better to have a lower DBI score as it signifies a more significant distinction between clusters [19]. After analysis, the data showed two distinct clusters when K=2.

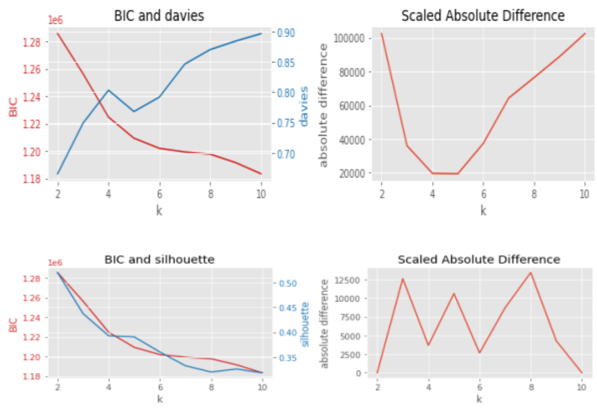

Fig. 9: Unsupervised evaluation metrics

C LSTM Results and Analysis of Models' Performance

As presented in Table V, various deep learning models have been evaluated to assess their potential applications, such as LSTM, BiLSTM, Stacked LSTM, and CNN+LSTM.

Table III and Fig. 10 analysis suggest that the CNN+LSTM model with a sequence length of 5 is the most effective, with a high F1 score of 0.91. The stacked LSTM model follows with a score of 0.70, and the LSTM model comes in third with a score of 0.58. The F1 score is an important metric to consider when dealing with imbalanced datasets.

Based on the results of the model testing, it was found that BiLSTM had poor performance in ROC_AUC as compared to the other models when tested at sequence lengths of 5 and 10. However, it showed better performance when tested at a sequence length of 15. This suggests that BiLSTM may perform better when given sufficient time to learn the true pattern of the system. Therefore, it may be recommended to use BiLSTM for higher sequence lengths in order to improve its performance. Stacked LSTM had the highest f1 score, followed by Stacked LSTM, and BiLSTM had the worst performance at the sequence length of 5. ROC AUC is a widely used evaluation metric for classification tasks, measuring a model's ability to distinguish between positive and negative instances.

Table III: LSTM Model Evaluation Metrics Results

Model Architectures	Seq_Length	Trainable Parameters	Recall	Precision	F1	ROC_AUC	Accuracy
LSTM	5	43 701	0,435	0,892	0,585	0,93	0,986
LSTM(Units=100) Dropout=0.80 Optimizer=SGD Loss=Binary crossentropy	10		0,757	0,466	0,577	0,94	0,976
	15		0,103	0,91	0,185	0,89	0,98
BiLSTM	5	162 565	0,191	1	0,322	0,92	0,982
Bidirectional(units=276) Dropout=0.5 Optimizer=SGD Loss=Binary crossentropy	10		0,125	1	0,222	0,85	0,981
	15		0,872	0,424	0,571	0,94	0,972
Stacked LSTM	5	162 565	0,588	0,869	0,702	0,93	0,989
LSTM(units=100) LSTM(units=50) Dropout=0.5 Optimizer=SGD Loss=Binary Crossentropy	10		0,786	0,623	0,695	0,94	0,985
	15		0,781	0,624	0,694	0,95	0,985
CNN+LSTM	5	87 001	**0,915**	**0,905**	**0,909**	**0,97**	**0,996**
Conv1D(Filters=100,Kernel=8,Activation=relu Max pooling (pool size=2) LSTM(Units=100) Dropout=0.80 Optimizer=Adam Loss= Binary Crossentropy	10		0,877	0,355	0,505	0,96	0,963
	15		0,403	0,933	0,563	0,96	0,986

The CNN+LSTM model with sequence length 5 is the best model, followed by stacked LSTM. BiLSTM performs better at higher sequence lengths. Choosing the right evaluation metric is critical to measure model effectiveness when dealing with imbalanced datasets accurately.

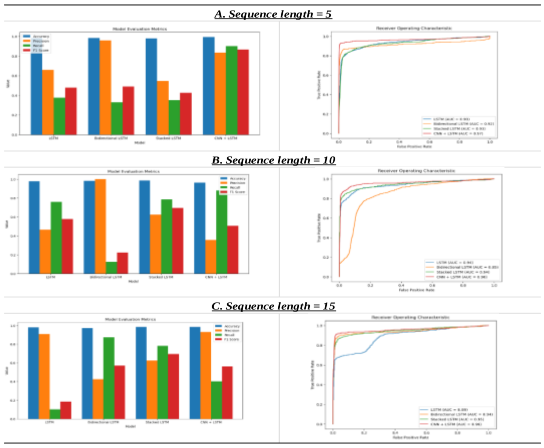

Fig. 10: Evaluation results of different models at different sequence lengths

IV. DISCUSSION

To test the effectiveness of an LSTM model in preventing load losses, suppose the coal mill experienced a trip due to low PA flow at exactly 10:20 PM on February 22nd, 2021. Prior to this, an unsafe operation was recorded by the GMM model at 21:50 PM. Similar events occurred on February 19th and February 21st. These incidents were used to test LSTM models' ability to predict and warn against unsafe operations (Fig. 11).

Different LSTM architectures, including BiLSTM, Stacked LSTMs, and CNN+LSTM, were analyzed to

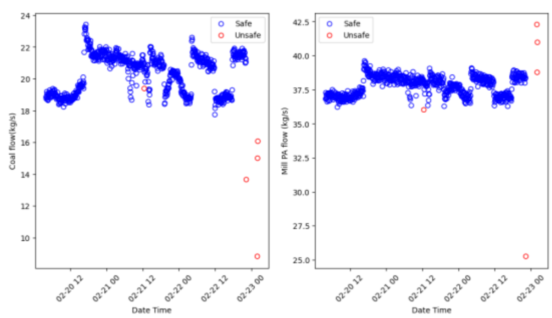

Fig. 11: Selected test scenario of Coal flow(kg/s) and Mill PA flow(kg/s) over he period of 4 days in 12 hours intervals

predict failure. The CNN+LSTM model stood out for its sensitivity to detecting impending failure, as shown by a rise in output values before the moment of failure on February 21st at 10:20 PM (Fig. 12). This model's CNN component helped identify spatial patterns, while the

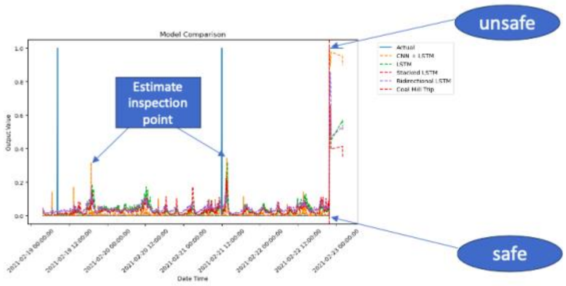

Fig. 12: This is a comparison of the values output by different LSTM models for the Coal Mill flow over a 5-day period, in 12-hour intervals. The "Actual" values refer to the GMM labels.

following LSTM layers monitored time-based relationships. In contrast, other LSTM architectures did not demonstrate the same sensitivity. The rise in output values of the CNN+LSTM model over time before the moment of failure aligns with expectations and underscores its potential for early detection of coal mill tripping events. The GMM+CNN+LSTM architecture was found to optimize failure detection by integrating temporal information from sensor data. The GMM output value indicates that the operation is highly unsafe, whereas the CNN+LSTM model estimates the output value around the GMM value estimate. Based on this, an inspection is estimated on February 21st , which is 75 minutes after the extreme GMM unsafe operation label . Overall, the study showed that the CNN+LSTM model was better suited for anticipating impending failure and may be a useful tool for early detection of similar events.

V. CONCLUSION

This study analyzed different LSTM models, including CNN+LSTM, to detect safe and unsafe operations of a coal mill using GMM for clustering. The models showed potential in accurately predicting unsafe modes, but there are limitations in generalizing to other coal mills or domains. Thus, it is necessary to fine-tune and optimize the models for practical deployment, but the analysis provided valuable insights for improving predictive maintenance strategies in the Power Engineering industry.

ACKNOWLEDGEMENT

The author wants to thank everyone who helped complete this dissertation. Firstly, the author is deeply grateful to Dr. Akihiro Fujii for his guidance and support throughout the research process. Ms. Kun Xiang also provided invaluable assistance and advice. Mr. Naoya Tanaka's insights added significant value to the author's research. The author also acknowledges the support of Eskom Holding SOC Ltd and the Japan International Cooperation Agency (JICA). The author's husband, Pitso Sekhoto, reviewed the dissertation and provided constant love and support. Lastly, the author thanks her family for their unwavering encouragement. The author is forever grateful to everyone who contributed to her academic journey.

REFERENCES

[1] Y. Hu, B. Ping, D. Zeng, Y. Niu and Y. Gao, "Modeling of Coal Mill System Used for Fault Simulation," *Energies,* vol. 13, no. 7, p. 1784, April 2020.

[2] X. Li, Y. Wu, H. Chen, Y. Zhou, L. Wu, K. Chen and K. Cen, "Coal mill model considering heat transfer effect on mass equations with estimations of moisture," *Journal of Process Control,* vol. 104, pp. 178-188, 2021.

[3] P. F. Odgaard, B. Lin and S. B. Jørgensen, "Observer - Based and Regression Model-Based Detection of Emerging Faults in Coal Mills," *Fault Detection, Supervision, and Safety of technical Processes,* vol. 23, no. 2, pp. 659-668, June 2008.

[4] M. Little, "Impact of Pulverizer Performance' Pulverizer Diagnostics for improved Performance and Reliability," Engineering Consultants Group, CA, 2014.

[5] E. Das and H. Pannen, "Description of Grinding Plant," Hitachi, technical report: B1141116-35-99-IB07-00001-AE, 2009.

[6] V. Agrawal, B. K. Panigrahi and P. M. V. Subbarao, "Review of control and fault diagnosis methods applied to coal mills," *Process Control,* vol. 32, no. 0959-1524, pp. 138-153, 2015.

[7] P. Odgaard, B. Lin and S. B. Jørgensen, "Observer and Data-Driven-Model-Based Fault Detection in Power Plant Coal Mills," *IEEE transactions on Energy Conversion*, vol. 23, no. 2, pp. 659-668, 2008.

[8] Y. G. Zhang, Q. H. Wu, G. Oluwande, D. Matts and X. Zhou, "Coal mill modelling by machine learning based on onsite measurements," *IEEE Transactions on Energy Conversion*, vol. 17, no. 4, pp. 549-555, 2002.

[9] J. Wei, J. Wang, and S. Guo, "Mathematics modeling and condition monitoring of power station ball mill," in *2009 American Control Conference*, 2009.

[10] W. Dayong, Y. Changwei, W. Kumfer and L. Hongchao, "A life-cycle optimization model using semi-Markov process for highway bridge maintenance," *Applied Mathematical Modelling,* vol. 43, pp. 45-60, 2017.

[11] Y. Jian, X. Qing, Y. Zhao, L. He and X. Qi, "Application of Model-Based Deep Learning Algorithm in Fault Diagnosis of Coal," *Mathematical Problems in Engineering,* p. 14, August 2020.

[12] H. Pariaman, GM Luciana, M.K Wisyaldin, and M Hisjam, "Anomaly Detection Using LSTM-Autoencoder to Predict Coal Pulverizer Condition on Coal-Fired Power Plant." in *Evergreen*, ISSN: 2189-0420, vol.8, no.1, pp 89-97

[13] H. Zhang, C. Pan, Y. Wang, M. Xu, F. Zhou, X. Yang, L. Zhu, C. Zhao, Y. Song, and H. Chen," Fault Diagnosis of Coal Mill Based on Kernel Extreme

Learning Machine with Variational Model Feature Extraction” in *Energies*, 2022.

[14] X. Hong, Z. Xu and Z. Zhang, "Abnormal Condition Monitoring and Diagnosis for Coal Mills Based on Support Vector Regression," in *IEEE Access*, vol. 7, pp. 170488-170499, 2019, DOI: 10.1109/ACCESS.2019.2955249.

[15] W. Fan, S. Ren, Q. Zhu, Z. Jia, D. Bai, and F. Si, "A Novel Multi-Mode Bayesian Method for the Process Monitoring and Fault Diagnosis of Coal Mills," in *IEEE Access*, vol. 9, pp. 22914-22926, 2021, DOI: 10.1109/ACCESS.2021.3055226.

[16] Géron, A, "Gaussian Mixtures," in *Hands on machine Learning with Sci-kit Learn. Keras, Tensorflow*, CA, O" Reily Media Inc., 2019, p. 239.

[17] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian Inference based finite Gaussian mixture models," *AIChe J*, vol. 54, no. 7, pp. 188-1929, May 2008.

[18] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381-396, March 2002.

[19] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vols. PAMI-1, no. 2, pp. 224-227, April 1979.

[20] A. K. Singh, S. Mittal, P. Malhotra and Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index (DBI) in Cereal data using Kmeans," in *Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020.

[21] J. Davis and M. Goadrich, "The relationship Between Precision-Recall and ROC Curves," in *proceedings of the 23rd International Conference on Machine learning (ICML '06)*, Pittsburg, USA, 2006.

[22] Powers D. M. W. "Evaluation: from precision, recall and F1 measure to ROC, informedness, markedness and correlation," *Journal of Machine learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.