Predicting Patient Triage at the Emergency Department using Machine Learning Classification: The Case Study of UNS Hospital, Indonesia

Pringgo Widyo Laksono
Industrial Engineering, Universitas Sebelas Maret, Surakarta, Indonesia
pringgo@ft.uns.ac.id (corresponding author)

Rizki Ananda Putra Nur Rohmat
Industrial Engineering, Universitas Sebelas Maret, Surakarta, Indonesia
rizkianandapnr@gmail.com

Retno Wulan Damayanti
Industrial Engineering, Universitas Sebelas Maret, Surakarta, Indonesia
rwd@ft.uns.ac.id

Eko Pujiyanto
Industrial Engineering, Universitas Sebelas Maret, Surakarta, Indonesia
ekopujiyanto@ft.uns.ac.id

Cucuk Nur Rosyidi
Industrial Engineering, Universitas Sebelas Maret, Surakarta, Indonesia
cucuknur@staff.uns.ac.id

Received: 16 March 2025 | Revised: 16 April 2025 | Accepted: 19 April 2025

ABSTRACT

The triage of patients in the Emergency Department (ED) plays a critical role in determining the urgency and type of treatment that shall be administered. Therefore, an accurate prediction system for patient triage can be very helpful. This study aims to develop a Machine Learning (ML) classification model to predict triage decisions for patients admitted to the ED at UNS Hospital. Several classification models were evaluated, including Naïve Bayes (NB), Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN). To assess model performance, multiple metrics were employed: accuracy, precision, recall, F1-score, and the confusion matrix. Following an initial evaluation, hyperparameter tuning was conducted on the selected best-performing model to identify the optimal combination of parameters for improved predictive performance. RF emerged as the best-performing model for triage prediction. The parameters tuned included the number of estimators, criterion, maximum features, and maximum depth, using a 5-fold cross-validation strategy. The optimal parameter values were found to be 900 estimators, 'entropy' for the criterion, 'log2' for max features, and a maximum depth of 40. The results of this study demonstrate that hyperparameter tuning can significantly enhance model performance, reducing recall errors, improving the F1-score, and decreasing the number of mispredictions.

Keywords-emergency department; triage; machine learning; classification

I. INTRODUCTION

One of the primary objectives of a hospital is to provide emergency care to patients. To fulfill this need, every hospital is required to operate an Emergency Department (ED), whose main goal is to deliver immediate treatment for illnesses or injuries that could threaten patients' lives [1]. This critical role emphasizes the importance of effective treatment management for patients admitted to the ED. Patient triage forms the foundation of ED operations. Triage involves classifying patients into several categories, based on patients' conditions, prognoses, and resource availability [2, 3], to determine which patients should receive priority medical treatment and in which department to allocate them. At the ED of UNS Hospital, patients are categorized into five triage levels, with level 1 representing those requiring immediate treatment and level 5 for patients with the lowest urgency. Typically, triage assessment is performed as soon as the patient arrives at the ED entrance manually by the triage medical personnel, and this is the case for UNS Hospital. However, several challenges arise from this manual process, such as delays in triage assessment due to the absence of personnel at critical times and errors in assigning patients to the appropriate triage categories.

Intelligent technologies like the Internet of Things (IoT), Machine Learning (ML), and Artificial Intelligence (AI) are integrated to enhance system usability and performance [4] and offer benefits to the healthcare system. For example, Amiens Picardy University Hospital has developed AI models to predict patient outcomes and optimize resource usage, including time and hospital bed management [5]. Unsupervised ML techniques have been utilized to assess patient severity, intubation needs, length of stay, and oxygen saturation, aiding critical decision-making [6]. Similarly, ML models have been applied to support cardiovascular prognosis decision-making in the ED [7], to categorize mental health conditions [8], and to forecast pediatric ED admissions using Deep Neural Networks (DNN) [9]. Meanwhile, [10] uses ML clustering analysis to ascertain whether a new triage method can safely send certain ED patients to a General Practitioner Cooperative (GPC). It compares several ML classification models, such as Naïve Bayes (NB), Decision Trees (DTs), Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbors (k – NN). The hypothetical data are taken from [11].

What distinguishes this research is its focus on applying ML techniques to real-world triage data from a public hospital in Indonesia. This dataset includes operational and culturally specific variables that have been largely overlooked in previous studies. While earlier research has primarily focused on ML applications in cardiovascular triage or ED referral decision-making [7, 10], our study broadens the scope by applying classification models to a wider patient population and by incorporating localized features and contextual triage outcomes. This study compared several classification models used in [7-9] to predict patient triage at the ED of the UNS Hospital. Furthermore, hyperparameter tuning was conducted on the selected model to determine the optimal parameters for improving triage prediction performance. The final model is expected to automate the triage process, accelerating patient care and supporting the workload of medical personnel, particularly within the ED of UNS Hospital.

II. MATERIALS AND METHODS

The model development process in this research consisted of several key stages: data preprocessing, dataset splitting, data encoding, data rescaling, model construction, model evaluation, and hyperparameter tuning. The overall workflow for creating the classification model is illustrated in Figure 1.

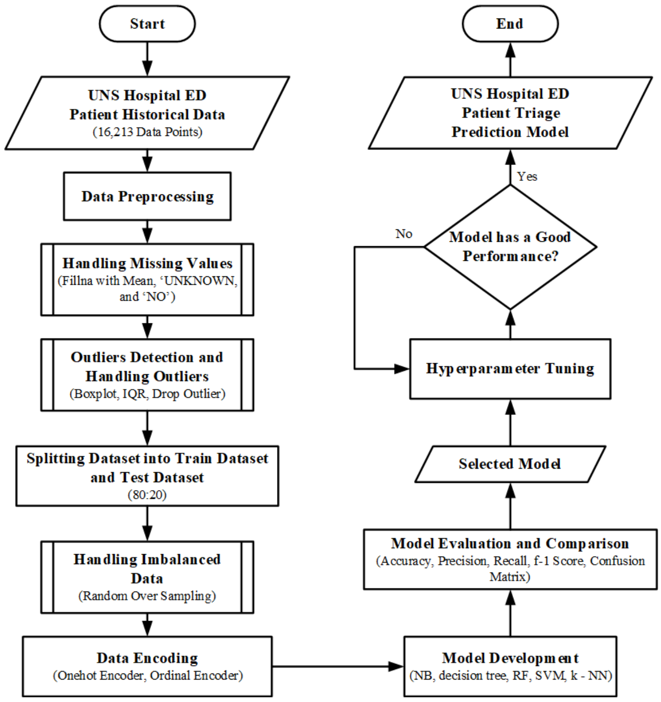

Fig. 1. ML model architecture.

A. Data

This research investigated the medical records of patients at the ED of UNS Hospital in 2019. From these record, this study examined 16,213 data with several features including: *GENDER, AGE, HOSPITALIZATION, REFERRED?, DISCHARGED?, DOR?, DOA?, CARDIAC_ARREST?, CHEST_ PAIN?, ABDOMINAL_ PAIN?, EPIGASTRIC _PAIN?, SEIZURE?, FEVER?, ASTHMA?, MCU?, HEAD INJURY, DECREASED CONSCIOUSNESS?, CHRONIC KIDNEY DISEASE?, TRAFFIC ACCIDENT?, VERTIGO?, VL?, VE?, HT?, VOMITUS?, DM?, CASE, SPECIALIST CONSULTATION?*, and *TRIAGE*. The descriptive statistics of numerical and categorical data are presented in Tables I and II.

Table II summarizes the descriptive statistics for categorical variables in the dataset. The "Unique" column indicates the number of distinct categories for each feature. The "Top" column shows the most frequent category, and "Freq." represents how many times that top category appears. The "% of Total" column shows the proportion of the top category relative to the total number of records (16,213). The "count" column has been omitted since all features share the same total.

TABLE I. DESCRIPTIVE STATISTICS OF NUMERICAL DATA

	count	mean	st.d	min	25%	50%	75%	max
AGE	16,213	34.753415	23.156746	0.0	18.0	30.0	55.0	287.0
TRIAGE	16,213	4.052735	0.625843	1.0	4.0	4.0	4.0	5.0

st.d: standard deviation

TABLE II. DESCRIPTIVE STATISTICS OF CATEGORICAL DATA

	Unique	Top	Freq.	% of Total
GENDER	3	P	8,142	50.2
HOSPITALIZATION	2	YES	12,769	78.7
REFERRED?	2	NO	16,066	99.1
DISCHARGED?	2	NO	12,948	79.9
DOR?	2	NO	16,102	99.3
DOA?	2	NO	16,195	99.9
CARDIAC_ARREST?	2	NO	16,138	99.5
CHEST_PAIN?	2	NO	16,014	98.8
ABDOMINAL_PAIN?	2	NO	15,284	94.3
EPIGASTRIC_PAIN?	2	NO	16,174	99.8
SEIZURE?	2	NO	15,978	98.6
FEVER?	2	NO	12,475	76.9
ASTHMA?	2	NO	15,906	98.1
MCU?	2	NO	15,734	97.0
HEAD INJURY	4	NO	15,810	97.5
DECREASED CONSCIOUSNESS?	2	NO	16,015	98.8
CHRONIC_KIDNEY_DISEASE?	2	NO	15,987	98.6
TRAFFIC ACCIDENT?	2	NO	16,010	98.8
VERTIGO?	2	NO	15,756	97.2
VL?	2	NO	15,764	97.2
VE?	2	NO	15,962	98.5
HT?	2	NO	15,657	96.6
VOMITUS?	2	NO	15,306	94.4
DM?	2	NO	15,795	97.4
CASE	23	NO	7,786	48.0
SPECIALIST CONSULTATION?	2	NO	12,235	75.5

B. Methods

1) Data Preprocessing

The data preprocessing stage involved several key steps, including the treatment of missing values and outliers. Instead of discarding any columns, specific imputation strategies were applied. Missing values in the AGE variable were replaced with the mean of the available values. For categorical variables such as CASE and GENDER, missing values were filled with "UNKNOWN". For binary categorical features, missing values were replaced with "NO". Outliers were addressed using the Interquartile Range (IQR) method [13, 14]. A boxplot was used to visualize the distribution, revealing only two extreme outliers in the AGE variable. These entries were removed to preserve data integrity.

2) Splitting the Dataset

Compared to other sources of model uncertainty, such as network initialization or training dynamics, the uncertainty due to sample variance can be substantially larger [15]. In this study, 80% of the dataset was used for the training of the model, and the rest 20% was used as the test set for model validation. The train-test-split function from Scikit-learn was utilized for this purpose. After splitting, data imbalance was addressed using an oversampling approach to balance class distributions and improve model performance [16]. Figure 2 shows the split dataset before and after handling the imbalanced data.

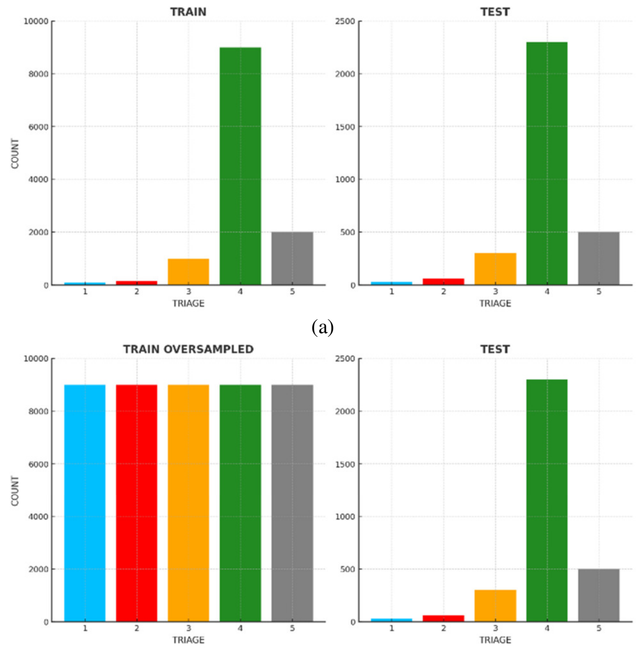

Fig. 2. (a) Train dataset and test dataset before handling imbalanced data, and (b) train dataset and test dataset after handling imbalanced data.

3) Data Encoding

Data encoding is the process of converting categorical features into numerical representations, which is essential for ML algorithms. The choice of encoding method can influence the expressiveness of variational ML models [17]. In this study, the OrdinalEncoder was used for ordinal categorical features, while the OneHotEncoder was applied to nominal categorical features.

4) Model Development

The prediction models developed in this study were based on various ML algorithms, including NB, DT, RF, SVM, and k-NN. During the initial model development, hyperparameter tuning was not performed; instead, each model was built using default parameters to allow for a fair baseline comparison.

5) Model Evaluation and Comparison

Model performance was evaluated using several metrics, including accuracy, precision, recall, F1-score, and the misclassification counts from the confusion matrix.

6) Hyperparameter Tuning

Hyperparameters significantly influence the predictive performance of ML models [18]. In this study, the GridSearchCV function from Scikit-learn was used to systematically search for the best hyperparameter configuration for the selected model based on the highest accuracy score obtained during cross-validation [19]. The tuning process involved evaluating 60 different parameter combinations of n_estimators, criterion, max_features, and max_depth, with a cross-validation value set to 5.

III. RESULTS AND DISCUSSION

Table III summarizes the average performance metrics of the models across the five triage levels. Figure 3 illustrates the confusion matrix, where True Positives (TP) appear along the diagonal from the top-left to bottom-right, while False Positives (FP) and False Negatives (FN) indicate misclassifications.

TABLE III. AVERAGE PERFORMANCE METRICS OF ML PREDICTION MODELS

ML Model	Accuracy	Precision	Recall	F1-score	Misclassification counts
NB	0.43	0.45	0.40	0.30	1837
DT	0.88	0.58	0.62	0.60	375
RF	0.89	0.60	0.64	0.62	363
SVM	0.83	0.60	0.66	0.61	554
k – NN	0.84	0.55	0.60	0.56	505

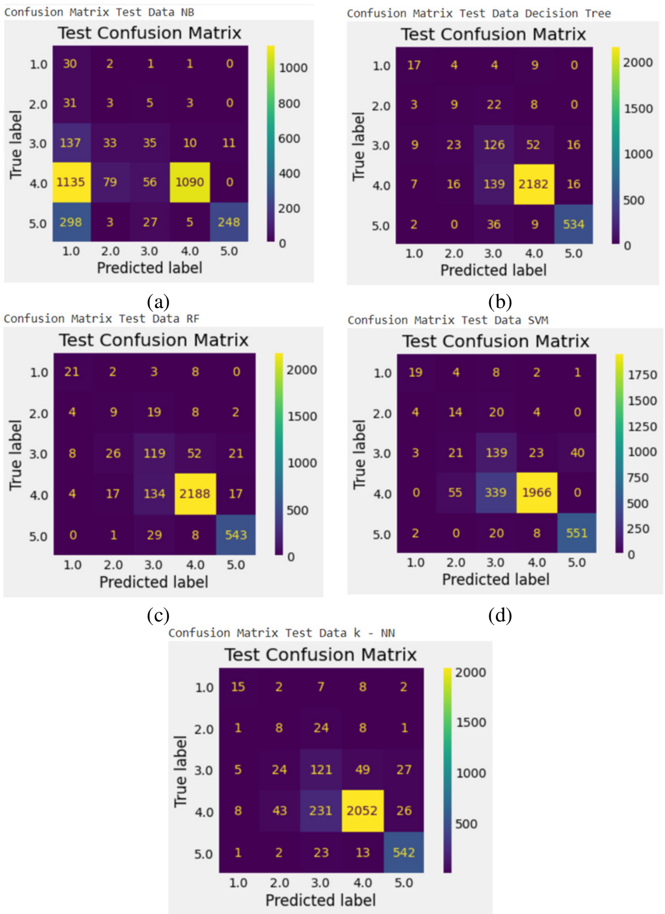

Fig. 3. Confusion matrix on test dataset for ML models: (a) NB, (b) DT, (c) RF, (d) SVM, and (e) k-NN.

TABLE IV. OPTIMAL HYPERPARAMETERS FOR THE RF MODEL

Hyperparameter	Values
N_estimators	900
Criterion	entropy
Max_features	Log2
Max_depth	40
Accuracy	0.98

As shown in Table III and Figure 3, the RF model outperformed the others, followed by DT and SVM, achieving an accuracy of 89%, an F1-score of 62%, and recording 363 misclassifications, corresponding to 11% of the test cases. Lower performance metrics, such as reduced accuracy, precision, recall, F1-score, and increased misclassifications, are attributed to data quality issues, including errors during data entry and inaccuracies in triage labeling. Furthermore, the limited number of input features restricted the model's ability to capture key patient characteristics. As the best-performing model, RF was selected for further tuning to optimize its hyperparameters. The optimal values identified for each hyperparameter are presented in Table IV. The RF model, updated with these parameters, was re-evaluated, with the post-tuning performance metrics reported in Table V and the final confusion matrix displayed in Figure 4.

TABLE V. PERFORMANCE MEASURE OF PREDICTION MODEL AFTER HYPERPARAMETER TUNING

Triage	Accuracy	Precision	Recall	F1-score	Misclassification counts
1	89%	0.54	0.56	0.55	356
2		0.15	0.19	0.17	
3		0.40	0.54	0.46	
4		0.97	0.93	0.95	
5		0.93	0.93	0.93	
Average		0.60	0.63	0.61	

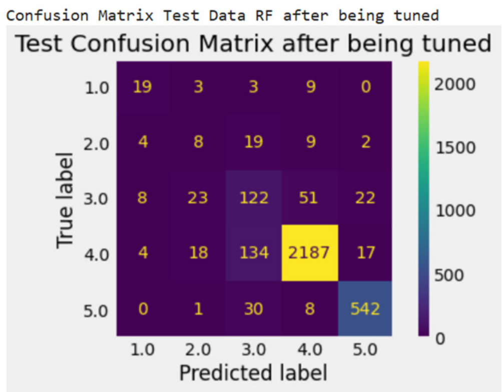

Fig. 4. RF model confusion matrix on selected test dataset after hyperparameter tuning.

Post-tuning, the recall score slightly decreased from 0.64 to 0.63, and the F1-score declined from 0.62 to 0.61. However, the number of misclassifications in the confusion matrix was reduced from 363 to 356.

IV. CONCLUSION

This study compares several Machine Learning (ML) classification models, including Naïve Bayes (NB), Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN), in their ability to classify patient triage level (1-5) based on severity and urgency in at the Emergency Department (ED) of UNS Hospital.

The RF model achieved the highest performance metrics, achieving an accuracy of 88%, with a precision of 0.60, a recall of 0.64, and an F1-score of 0.62, and produced the fewest false positives and false negatives, misclassifying 363 cases. Based on this performance, the RF model was selected for further optimization through hyperparameter tuning. The optimal values identified were 900 (n_estimators), entropy (criterion), log2 (max_features), and 40 (max_depth). After tuning, accuracy and precision remained unchanged, while recall slightly decreased to 0.63 and F1-score to 0.61. However, the number of misclassified cases was reduced from 363 to 356.

This research provides a comprehensive evaluation of ML models for triage prediction, based on medical records of patients at the ED of UNS Hospital in 2019. It highlights the potential of automated triage systems in resource-constrained settings—an area still evolving in current literature. Moreover, it contributes to the field of health informatics and supports practical efforts to enhance emergency care workflows and clinical decision-making.

ACKNOWLEDGMENT

This research is funded by the Research Group Grant of Universitas Sebelas Maret. The authors would like to express their sincere gratitude to UNS Hospital for providing access to patient medical records, which were essential for this study.

REFERENCES

[1] "Mengenal Instalasi Gawat Darurat (IGD) dan Pediatric Intensive Care Unit (PICU) di Rumah Sakit," *IDAI*, Feb. 10, 2017. https://www.idai.or.id/artikel/seputar-kesehatan-anak/mengenal-instalasi-gawat-darurat-igd-dan-pediatric-intensive-care-unit-picu-di-rumah-sakit.

[2] R. E. Wolfe, "Chapter 4 - Role of Emergency Medicine in Disaster Management," in *Ciottone's Disaster Medicine (Second Edition)*, G. R. Ciottone, Ed. Philadelphia: Elsevier, 2016, pp. 20–26, https://doi.org/10.1016/B978-0-323-28665-7.00004-2.

[3] K. B. Ahsan, M. A. Karim, G. J. FitzGerald, D. G. Morel, and J. A. Burke, "Development of Relationship between Triaging of Patients and Emergency Department Performance," *Procedia Manufacturing*, vol. 30, pp. 200–207, 2019, https://doi.org/10.1016/j.promfg.2019.02.029.

[4] G. Büchi, M. Cugno, and R. Castagnoli, "Smart factory performance and Industry 4.0," *Technological Forecasting and Social Change*, vol. 150, Jan. 2020, Art. no. 119790, https://doi.org/10.1016/j.techfore.2019.119790.

[5] E. Arnaud, M. Elbattah, C. Ammirati, G. Dequen, and D. A. Ghazali, "Use of Artificial Intelligence to Manage Patient Flow in Emergency Department during the COVID-19 Pandemic: A Prospective, Single-Center Study," *International Journal of Environmental Research and Public Health*, vol. 19, no. 15, Aug. 2022, Art. no. 9667, https://doi.org/10.3390/ijerph19159667.

[6] S. Boussen *et al.*, "Triage and monitoring of COVID-19 patients in intensive care using unsupervised machine learning," *Computers in Biology and Medicine*, vol. 142, Mar. 2022, Art. no. 105192, https://doi.org/10.1016/j.compbiomed.2021.105192.

[7] H. Jiang *et al.*, "Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease," *International Journal of Medical Informatics*, vol. 145, Jan. 2021, Art. no. 104326, https://doi.org/10.1016/j.ijmedinf.2020.104326.

[8] B. H. Bhavani and N. C. Naveen, "An Approach to Determine and Categorize Mental Health Condition using Machine Learning and Deep Learning Models," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13780–13786, Apr. 2024, https://doi.org/10.48084/etasr.7162.

[9] B. P. Roquette, H. Nagano, E. C. Marujo, and A. C. Maiorano, "Prediction of admission in pediatric emergency department with deep neural networks and triage textual data," *Neural Networks*, vol. 126, pp. 170–177, Jun. 2020, https://doi.org/10.1016/j.neunet.2020.03.012.

[10] S. Morreel *et al.*, "Triaging and referring in adjacent general and emergency departments (the TRIAGE trial): A cluster randomised controlled trial," *PLOS ONE*, vol. 16, no. 11, Nov. 2021, Art. no. e0258561, https://doi.org/10.1371/journal.pone.0258561.

[11] F. Gao, B. Boukebous, M. Pozzar, E. Alaoui, B. Sano, and S. Bayat-Makoei, "Predictive Models for Emergency Department Triage using Machine Learning: A Systematic Review," *Obstetrics and Gynecology Research*, vol. 05, no. 02, 2022, https://doi.org/10.26502/ogr085.

[12] O. H. Salman, Z. Taha, M. Q. Alsabah, Y. S. Hussein, A. S. Mohammed, and M. Aal-Nouman, "A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work," *Computer Methods and Programs in Biomedicine*, vol. 209, Sep. 2021, Art. no. 106357, https://doi.org/10.1016/j.cmpb.2021.106357.

[13] V. Kalra and R. Aggarwal, "Importance of Text Data Preprocessing & Implementation in RapidMiner," presented at the The First International Conference on Information Technology and Knowledge Management, Jan. 2018, pp. 71–75, https://doi.org/10.15439/2017KM46.

[14] S. Walfish, "A review of statistical outlier methods," *Pharmaceutical Technology*, vol. 30, no. 11, pp. 82–86, Nov. 2006.

[15] R. J. May, H. R. Maier, and G. C. Dandy, "Data splitting for artificial neural networks using SOM-based stratified sampling," *Neural Networks*, vol. 23, no. 2, pp. 283–294, Mar. 2010, https://doi.org/10.1016/j.neunet.2009.11.009.

[16] S. Mishra, "Handling Imbalanced Data : SMOTE vs. Random Undersampling," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 8, pp. 317–320, Aug. 2017.

[17] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Physical Review A*, vol. 103, no. 3, Mar. 2021, Art. no. 032430, https://doi.org/10.1103/PhysRevA.103.032430.

[18] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl, and A. C. P. L. F. Carvalho, "To tune or not to tune: Recommending when to adjust SVM hyper-parameters via meta-learning," in *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, Ireland, Jul. 2015, pp. 1–8, https://doi.org/10.1109/IJCNN.2015.7280644.

[19] G. S. K. Ranjan, A. Kumar Verma, and S. Radhika, "K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Bombay, India, Mar. 2019, pp. 1–5, https://doi.org/10.1109/I2CT45611.2019.9033691.