PT. Lentera Ilmu Publisher

Online version at https://journal.lenterailmu.com/index.php/josapen

JOSAPEN

JOURNAL OF COMPUTER SCIENCE APPLICATION AND ENGINEERING

E-ISSN: 0000-0000, P-ISSN: 0000-0000, DOI:-

First Step for Vehicle License Plate Identification Using Machine Learning Approach

Amirah[a]*, Ahmad Sanmorino[b]

[a]*Lentera Ilmu Publisher, Prabumulih, Indonesia*
[b]*Faculty of Computer and Science, Universitas Indo Global Mandiri, Palembang, Indonesia*

ARTICLE INFO

Article history:
Received 25 October 2022
Revised 15 December 2022
Accepted 06 January 2023

Keywords:
Machine learning
License plate identification
Neural networks
Transportation optimization
Ethical data handling

ABSTRACT

Automated vehicle license plate identification, critical in modern transportation systems, finds application in traffic monitoring, law enforcement, and transportation optimization. This study explores machine learning's potential to enhance accuracy and efficiency in this domain. Leveraging neural networks and pattern recognition, it aims to build an automated system robust across diverse conditions. Addressing limitations in traditional methods, it focuses on adapting to lighting, angles, and image quality variations. The societal impact includes streamlining law enforcement and optimizing traffic flow, revolutionizing transportation and surveillance. Methodologies cover data collection, ethical considerations, preprocessing, feature extraction, model selection, and iterative refinement. Ethical data handling ensures privacy compliance. Feature extraction methods like HOG, LBP, CNNs, and color histograms capture crucial aspects for identification. Model selection spans SVMs, CNNs, decision trees, and ensemble methods, considering task complexity and dataset characteristics. This study evaluates machine learning's potential for revolutionizing license plate identification systems.

1. Introduction

Automated vehicle license plate identification has emerged as a critical facet of modern transportation systems, facilitating various applications in traffic monitoring, law enforcement, and intelligent transportation systems. The integration of machine learning techniques within this domain presents a promising avenue for enhancing the accuracy and efficiency of license plate identification processes [1]-[3]. This preliminary study delves into the application of machine learning approaches for the purpose of vehicle license plate identification, aiming to explore the feasibility and efficacy of such methods in real-world scenarios.

The ubiquity of cameras in traffic surveillance systems has generated vast amounts of visual data, making manual identification of license plates a time-consuming and error-prone task. Leveraging machine learning algorithms, particularly computer vision models, offers the potential to automate and

* *Corresponding author:* Amirah
E-mail address: white99pasific@gmail.com

streamline this process, enabling rapid and accurate identification of license plates from images or video feeds [4], [5]. By harnessing the power of neural networks and pattern recognition techniques, this study seeks to lay the groundwork for an efficient automated system capable of robustly identifying license plates across diverse environmental conditions.

One of the primary motivations behind this research is to address the challenges associated with conventional methods of license plate identification, which often struggle with variations in illumination, angles, and image quality. Machine learning models, when properly trained and optimized, possess the capability to adapt to such variations, thereby improving the overall accuracy and reliability of license plate recognition systems [6], [7]. This preliminary investigation aims to assess the performance and limitations of machine learning approaches in handling these challenges, paving the way for future advancements in this field.

Moreover, the potential societal impact of deploying robust machine learning-based license plate identification systems is substantial. Enhanced automation in this domain not only expedites law enforcement activities but also contributes to the efficiency of toll collection, parking management, and traffic flow optimization. By reducing human intervention and minimizing error rates, these systems hold promise in enhancing overall transportation system efficacy, ultimately leading to safer and more organized road networks.

In summary, this preliminary study embarks on a quest to evaluate the viability of employing machine learning approaches for vehicle license plate identification. By addressing the limitations of existing methods and harnessing the potential of advanced algorithms, this research endeavors to contribute to the development of efficient and reliable systems that can significantly impact various facets of modern transportation and surveillance.

2. Method

Figure 1 shows several key steps to perform vehicle license plate identification using a machine learning approach.

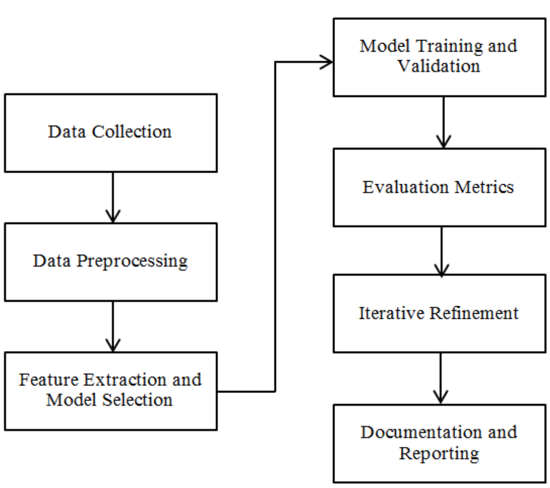

Figure 1 - Research method

2.1. Data Collection

Acquire a diverse dataset of vehicle images containing license plates. These images should encompass various lighting conditions, angles, weather scenarios, and types of vehicles. Manually annotate or label these images to indicate the location and content of the license plates. Ensure accuracy in the annotations as they will be used for training and validation.

2.2. Data Preprocessing

Resize, standardize, and preprocess the images to ensure consistency in format, resolution, and quality. This step might involve normalization, grayscale conversion, or other techniques to enhance the model's performance. Split the dataset into training, validation, and testing sets to train the machine learning model and assess its performance accurately.

2.3. Feature Extraction and Model Selection

Implement various computer vision techniques (e.g., edge detection, feature extraction) to extract relevant features from the license plate images. Choose and develop a suitable machine learning model architecture for license plate identification. Common choices include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or hybrid models tailored for object detection and recognition tasks [8], [9].

2.4. Model Training and Validation

Train the selected model using the labeled dataset. Fine-tune the model parameters and hyperparameters to optimize performance. Validate the trained model using the validation set to ensure it generalizes well and does not overfit the training data. Adjust the model as necessary based on validation performance.

2.5. Evaluation Metrics

Define evaluation metrics such as accuracy, precision, recall, and F1-score to quantitatively assess the model's performance in identifying license plates [10]. Evaluate the model on the test dataset to obtain final performance metrics and validate its real-world applicability.

2.6. Iterative Refinement

Analyze the model's performance and identify areas for improvement. This might involve data augmentation, fine-tuning model architecture, or experimenting with different preprocessing techniques. Iterate through model training, validation, and evaluation steps to refine the model and enhance its accuracy and robustness.

2.7. Documentation and Reporting

Document the entire methodology, including dataset details, preprocessing steps, model architecture, hyperparameters, and results obtained. Prepare a comprehensive report summarizing the findings, challenges faced, and potential avenues for further research and development in the field of vehicle license plate identification using machine learning.

3. Result and Discussion

In this preliminary study, we only discussed the third stage, namely Feature Extraction and Model Selection. The fourth stage and beyond will be carried out in future research.

3.1. Data Collection

There are several methods can consider for collecting data specifically for vehicle license plate identification using a machine learning approach (Table 1):

Table 1 – Data collecting methods

Methods	Kind of Methods	Description
Public Datasets	Open Datasets	Look for publicly available datasets related to vehicle images or license plates. Websites like Kaggle, GitHub, or government repositories may have relevant datasets [11].
	Traffic Cameras	Some cities or transportation departments release traffic camera images. These might contain images of vehicles and license plates.
	Research Repositories	Academic institutions or research organizations might have datasets available for specific purposes.
Collaboration and Partnerships	Local Authorities	Contact local government agencies, such as transportation departments or law enforcement, to inquire about access to their databases or the possibility of collaboration.
	Parking Lots or Garages	Collaborate with private parking facilities that might have surveillance cameras capturing vehicle images and license plates.
Crowdsourcing	Mobile Apps	Develop an app that allows users to voluntarily contribute images of vehicles or license plates. Ensure proper consent and data privacy measures.
	Online Platforms	Utilize crowdsourcing platforms where individuals can upload images anonymously [12].
Self-Captured Data	Capture Images	Use cameras, smartphones, or specialized equipment to capture images of vehicles and license plates in different conditions (lighting, weather, angles).
	Simulated Data	Create synthetic data that mimics real-world scenarios using software tools or simulation environments.
Web Scraping	Online Image Databases	Scrape publicly available vehicle image databases or websites with images of vehicles or license plates. Ensure compliance with terms of use and copyright regulations.
Sensor Data	IoT Devices	Utilize sensors or IoT devices attached to vehicles to capture images or data related to license plates [13].
Combination of Methods	Hybrid Approach	Combine multiple methods mentioned above to create a diverse and comprehensive dataset.

When collecting data always pay attention to Ethical and Legal Considerations. Always prioritize privacy and adhere to legal regulations regarding the collection and use of sensitive data, especially personally identifiable information like license plates. Anonymize or blur sensitive information in the collected images to protect privacy. Regardless of the method used, ensuring the quality, diversity, and ethical handling of the collected data is essential for training robust machine learning models. Table 2 shows an example of a small dataset for vehicle license plate identification. Each row represents an entry with features relevant to the study:

Table 2 – An example of a small dataset

Image_ID	License_Plate	Vehicle_Type	Color	Image_Path
1	ABC123	Sedan	Red	/path/to/image1.jpg
2	XYZ789	SUV	Blue	/path/to/image2.jpg
3	DEF456	Truck	White	/path/to/image3.jpg
4	GHI789	Sedan	Silver	/path/to/image4.jpg
5	JKL012	Hatchback	Black	/path/to/image5.jpg
6	MNO345	Sedan	Blue	/path/to/image6.jpg
7	PQR678	SUV	Red	/path/to/image7.jpg
8	STU901	Truck	White	/path/to/image8.jpg
9	VWX234	Sedan	Silver	/path/to/image9.jpg
10	YZA567	Hatchback	Yellow	/path/to/image10.jpg

- Image_ID: Unique identifier for each image in the dataset.
- License_Plate: The license plate number associated with the vehicle in the image.
- Vehicle_Type: Type of vehicle (e.g., Sedan, SUV, Truck, Hatchback).
- Color: Color of the vehicle.
- Image_Path: File path to the image in the dataset.

3.2. Data Preprocessing

Data preprocessing is a critical step in preparing dataset for training a machine learning model for vehicle license plate identification. Here are the steps (Table 3):

Table 3 – Data preprocessing steps

No	Steps	Kind of Steps	Description
1	Data Cleaning	Remove Duplicates	Check for and remove any duplicate images or entries in your dataset.
		Handle Missing Data	Address any missing or incomplete images or information. This might involve image restoration or data imputation techniques.
2	Image Preprocessing	Resizing and Standardizing	Resize images to a consistent size. Standardizing image dimensions helps ensure uniformity and efficient model training.
		Normalization	Normalize pixel values to a common scale (e.g., 0 to 1) to aid model convergence during training.
		Color Spaces	Convert images to appropriate color spaces (e.g., RGB, grayscale) based on the requirements of your model.
3	Noise Reduction and Enhancement	Noise Removal	Apply filters or techniques (e.g., Gaussian blur, median blur) to reduce noise or unwanted artifacts in images.
		Contrast and Brightness Adjustment	Enhance image quality by adjusting contrast, brightness, or sharpness as needed.
4	Data Augmentation	Image Augmentation	Generate additional training data by applying transformations such as rotation, flipping, scaling, or adding artificial noise. This helps improve model robustness.
		Augmentation Techniques	Utilize libraries like OpenCV or Augmentor to implement augmentation strategies.
5	Label Encoding	Encode Labels	Convert categorical labels (license plate numbers) into numerical or encoded formats that machine learning algorithms can process.
6	Splitting Dataset	Train-Validation-Test Split	Divide the dataset into separate sets for training, validation, and testing. Typical splits might include 70-80% for training, 10-15% for validation, and 10-15% for testing.
7	Data Balancing (if necessary)	Class Imbalance	Address any imbalance in the dataset, especially if certain classes (license plate types or patterns) are underrepresented. Techniques like oversampling, undersampling, or using class weights can help mitigate this issue.
8	Metadata Handling	Record Metadata	Keep track of metadata associated with images, including acquisition details, labeling methods, and any modifications made during preprocessing.

9	Ethical Considerations	Privacy Measures	Ensure that sensitive information, such as personally identifiable data on license plates, is appropriately handled or obscured to protect privacy.
10	Documentation	Create Documentation	Maintain detailed records of the preprocessing steps applied to the dataset for reproducibility and transparency in research.

Each of these steps is crucial in creating a clean, standardized, and balanced dataset that can enhance the performance of machine learning model for license plate identification. Adjust these steps according to the specifics of the dataset and the requirements of the machine learning algorithm.

3.3. Feature Extraction and Model Selection

In vehicle license plate identification using machine learning, various feature extraction methods can be employed. Here are some common techniques (Table 4):

Table 4 - Feature extraction techniques

Techniques	Description	Usage
Histogram of Oriented Gradients (HOG) [14]	Computes gradients' magnitude and orientation in localized portions of an image to capture shape and edge information.	HOG can be applied to extract features from the regions of interest in images containing license plates, highlighting the shape and structure of characters or patterns.
Local Binary Patterns (LBP) [15]	Examines the local texture patterns in an image by comparing each pixel with its neighboring pixels.	LBP can capture texture variations on the license plate, aiding in characterizing the surface patterns, which might be helpful for identification.
Convolutional Neural Networks	CNNs automatically learn features by employing convolutional layers, capturing hierarchical features from images.	Train CNNs to extract relevant features directly from the images, potentially learning complex representations that aid in license plate identification.
Color Histograms [16]	Represents the distribution of colors within an image or specific regions, providing insights into color-based features.	Extract color histograms from the license plate area, enabling identification based on color distributions.
Edge Detection [17]	Identifies edges and sharp changes in pixel intensity, crucial for determining boundaries and shapes.	Detect edges within the license plate area to outline characters or distinguish patterns.
Gabor Filters [18]	Filters sensitive to frequencies and orientations, suitable for capturing texture information.	Apply Gabor filters to extract texture features from the license plate, enhancing the model's ability to differentiate patterns.
Character Segmentation	Identifies and separates individual characters or elements within the license plate.	Segmentation helps in isolating characters, facilitating their individual identification and interpretation.

These feature extraction methods aim to capture various aspects of the images, such as shape, texture, color, and spatial information, which are vital for distinguishing and identifying license plates. Choosing the appropriate method or a combination thereof depends on the characteristics of the dataset and the specific requirements of the identification task.

For vehicle license plate identification using machine learning, selecting an appropriate model depends on several factors, including the complexity of the task, the size and nature of the dataset, and computational considerations. Here are some machine learning models commonly used for this type of task (See Table 5):

Table 5 – Machine learning model

Model	Description	Usage
Support Vector Machines (SVM) [19], [20]	SVMs are effective for binary classification tasks and work well with high-dimensional data. They find the best hyperplane that separates classes in the feature space.	SVMs can be employed for license plate identification, especially for binary classification tasks like recognizing whether a given image contains a license plate or not.
Convolutional	CNNs are powerful	CNNs can be

Neural Networks (CNNs) [21]	for image-related tasks, capable of learning complex hierarchical features from images.	utilized for end-to-end learning, directly taking in image data and learning representations for license plate identification.
Random Forests or Decision Trees [22]	Decision tree-based models like Random Forests are versatile and work well with structured data. They can handle non-linear relationships and interactions between features.	Effective for both classification and feature importance analysis in license plate identification tasks.
Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) [23]	RNNs, especially LSTM variants, are suitable for sequential data processing. They maintain memory of past inputs, beneficial for sequential tasks.	Helpful when there's a need to interpret sequences or patterns in license plate characters.
Ensemble Learning Methods [24]	Techniques like AdaBoost, Gradient Boosting, or Bagging involve combining multiple models to improve overall performance.	Ensemble methods can boost accuracy by leveraging the strength of different models for license plate identification.
K-Nearest Neighbors (KNN)	KNN is a simple and intuitive algorithm that classifies objects based on the majority vote of their neighbors.	Suitable for smaller datasets and can be effective when combined with appropriate feature representations.
Deep Learning Architectures (Autoencoders, Siamese Networks)	Architectures like autoencoders or Siamese networks are beneficial for learning robust representations or performing similarity-based tasks.	Useful when the emphasis is on feature learning or comparison tasks in license plate identification.

When determining a machine learning model, you need to consider:

- **Data Size:** Deep learning models might require larger datasets for effective training, whereas simpler models like SVM or decision trees might perform well with smaller datasets.
- **Complexity of Task:** If the task involves both detection and character recognition, a combination of models or a model designed for end-to-end processing might be necessary.

Selecting the right model involves experimenting with different architectures and techniques, considering the nature of the dataset and the specific requirements of the license plate identification task.

4. Conclusion

Automated license plate identification through machine learning stands as a pivotal tool in modern transportation systems. This study delves into employing machine learning to enhance the accuracy and efficiency of license plate identification, crucial for various applications in traffic monitoring, law enforcement, and transportation systems. The integration of machine learning tackles challenges faced by traditional methods, adapting to lighting variations, angles, and image quality. By leveraging neural networks and pattern recognition, this study aims to pave the way for an automated system capable of identifying license plates across diverse conditions.

Beyond addressing current limitations, this research explores societal impacts by streamlining law enforcement, optimizing traffic flow, and contributing to the efficiency of transportation systems. The study is a stepping stone towards developing efficient and reliable systems, poised to revolutionize modern transportation and surveillance. The methodology outlined encompasses data collection, preprocessing, feature extraction, model selection, and ongoing iterative refinement. Crucially, ethical considerations are paramount in data collection and handling, ensuring privacy and compliance with legal regulations. Feature extraction methods like HOG, LBP, CNNs, and color histograms aim to capture critical aspects such as shape, texture, and color, essential for identifying license plates. Model selection involves a spectrum of options like SVMs, CNNs, decision trees, and ensemble methods, balancing complexity with dataset characteristics.

This study’s foundation lies in evaluating machine learning's viability for license plate identification. It serves as a roadmap, striving towards refined systems that significantly impact transportation and surveillance. This conclusion summarizes the study's objectives, methodologies, ethical considerations, and the overarching aim to revolutionize license plate identification systems through machine learning.

Acknowledgements

We would like to acknowledge Lentera Ilmu Publisher for supporting this work.

REFERENCES

[1] K. B. Sathya, S. Vasuhi, and V. Vaidehi, “Perspective Vehicle License Plate Transformation using Deep Neural Network on Genesis of CPNet,” *Procedia Comput. Sci.*, vol. 171, pp. 1858–1867, 2020, doi: 10.1016/j.procs.2020.04.199.

[2] W. Puarungroj and N. Boonsirisumpun, “Thai License Plate Recognition Based on Deep Learning,” *Procedia Comput. Sci.*, vol. 135, pp. 214–221, 2018, doi: 10.1016/j.procs.2018.08.168.

[3] D. Neupane, A. Bhattarai, S. Aryal, M. R. Bouadjenek, U. Seok, and J. Seok, "Shine: A deep learning-based accessible parking management system," *Expert Syst. Appl.*, vol. 238, no. PF, p. 122205, 2024, doi: 10.1016/j.eswa.2023.122205.

[4] M. S. H. Onim *et al.*, "BLPnet: A new DNN model and Bengali OCR engine for Automatic Licence Plate Recognition," *Array*, vol. 15, no. March, p. 100244, 2022, doi: 10.1016/j.array.2022.100244.

[5] T. Panchal, H. Patel, and A. Panchal, "License Plate Detection Using Harris Corner and Character Segmentation by Integrated Approach from an Image," *Procedia Comput. Sci.*, vol. 79, pp. 419–425, 2016, doi: 10.1016/j.procs.2016.03.054.

[6] E. Bentafat, M. M. Rathore, and S. Bakiras, "Towards real-time privacy-preserving video surveillance," *Comput. Commun.*, vol. 180, no. September, pp. 97–108, 2021, doi: 10.1016/j.comcom.2021.09.009.

[7] L. Zheng, X. He, B. Samali, and L. T. Yang, "An algorithm for accuracy enhancement of license plate recognition," *J. Comput. Syst. Sci.*, vol. 79, no. 2, pp. 245–255, 2013, doi: 10.1016/j.jcss.2012.05.006.

[8] M. A. Jawale, P. William, A. B. Pawar, and N. Marriwala, "Measurement : Sensors Implementation of number plate detection system for vehicle registration using IOT and recognition using CNN," *Meas. Sensors*, vol. 27, no. May, p. 100761, 2023, doi: 10.1016/j.measen.2023.100761.

[9] I. Slimani, A. Zaarane, W. Al Okaishi, I. Atouf, and A. Hamdoun, "An automated license plate detection and recognition system based on wavelet decomposition and CNN," *Array*, vol. 8, no. May, p. 100040, 2020, doi: 10.1016/j.array.2020.100040.

[10] P. Fränti and R. Mariescu-Istodor, "Soft precision and recall," Pattern Recognit. Lett., vol. 167, pp. 115–121, 2023, doi: 10.1016/j.patrec.2023.02.005.

[11] G. Robles, M. R. V. Chaudron, R. Jolak, and R. Hebig, "A reflection on the impact of model mining from GitHub," *Inf. Softw. Technol.*, vol. 164, no. May, p. 107317, 2023, doi: 10.1016/j.infsof.2023.107317.

[12] D. Deichmann, T. Gillier, and M. Tonellato, "Getting on board with new ideas: An analysis of idea commitments on a crowdsourcing platform," *Res. Policy*, vol. 50, no. 9, p. 104320, 2021, doi: 10.1016/j.respol.2021.104320.

[13] M. M. Abdellatif, N. H. Elshabasy, A. E. Elashmawy, and M. AbdelRaheem, "A low cost IoT-based Arabic license plate recognition model for smart parking systems," *Ain Shams Eng. J.*, vol. 14, no. 6, p. 102178, 2023, doi: 10.1016/j.asej.2023.102178.

[14] O. El Melhaoui and S. Benchaou, "An Efficient Signature Recognition System Based on Gradient Features and Neural Network Classifier," *Procedia Comput. Sci.*, vol. 198, pp. 385–390, 2021, doi: 10.1016/j.procs.2021.12.258.

[15] T. Agbele, B. Ojeme, and R. Jiang, "Application of local binary patterns and cascade AdaBoost classifier for mice behavioural patterns detection and analysis," *Procedia Comput. Sci.*, vol. 159, pp. 1375–1386, 2019, doi: 10.1016/j.procs.2019.09.308.

[16] Y. Miao and M. Yamaguchi, "A point selection strategy with edge and line detection for Direct Sparse Visual Odometry," *Graph. Vis. Comput.*, vol. 6, p. 200051, 2022, doi: 10.1016/j.gvc.2022.200051.

[17] C. Zhou *et al.*, "Multi-scale pseudo labeling for unsupervised deep edge detection," *Knowledge-Based Syst.*, vol. 280, no. October, p. 111057, 2023, doi: 10.1016/j.knosys.2023.111057.

[18] M. T. T. Cho, A. Chueasamat, T. Hori, H. Saito, and Y. Kohgo, "Effectiveness of filter gabions against slope failure due to heavy rainfall," *Soils Found.*, vol. 61, no. 2, pp. 480–495, 2021, doi: 10.1016/j.sandf.2021.01.010.

[19] I. Zoppis, G. Mauri, and R. Dondi, Kernel methods: Support vector machines, vol. 1–3. *Elsevier Ltd.*, 2018. doi: 10.1016/B978-0-12-809633-8.20342-7.

[20] A. Gatera, M. Kuradusenge, G. Bajpai, C. Mikeka, and S. Shrivastava, "Comparison of random forest and support vector machine regression models for forecasting road accidents," *Sci. African*, vol. 21, p. e01739, 2023, doi: 10.1016/j.sciaf.2023.e01739.

[21] S. Belattar, O. Abdoun, and E. K. Haimoudi, "Performance analysis of the application of convolutional neural networks architectures in the agricultural diagnosis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 1, pp. 156–162, 2022, doi: 10.11591/ijeecs.v27.i1

[22] Y. Li, E. Herrera-Viedma, G. Kou, and J. A. Morente-Molinera, "Z-number-valued rule-based decision trees," *Inf. Sci. (Ny).*, vol. 643, no. May, p. 119252, 2023, doi: 10.1016/j.ins.2023.119252.

[23] M. W. Hasan, "Building an IoT temperature and humidity forecasting model based on long short-term memory (LSTM) with improved whale optimization algorithm," *Memories - Mater. Devices, Circuits Syst.*, vol. 6, no. October, p. 100086, 2023, doi: 10.1016/j.memori.2023.100086.

[24] L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. dos S. Coelho, "Ensemble learning methods using the Hodrick–Prescott filter for fault forecasting in insulators of the electrical power grids," *Int. J. Electr. Power Energy Syst.*, vol. 152, no. June, p. 109269, 2023, doi: 10.1016/j.ijepes.2023.109269.