BAREKENG: Journal of Mathematics and Its Applications
September 2025 Volume 19 Issue 3 Page 1597-1608
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss3pp1597-1608

TRANSFORMER-BASED OPTICAL CHARACTER RECOGNITION APPROACH FOR IDENTIFYING MOTOR VEHICLES WITH OVERDUE TAXES

Nabila Dwi Fazira[1], Achmad Fauzan[2*]

[1,2]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia
Jln. Kaliurang KM. 14,5 Sleman, Yogyakarta, 55584, Indonesia.

*Corresponding author's e-mail: * **achmadfauzan@uii.ac.id***

ABSTRACT

Article History:

Received: 26th December 2024
Revised: 22nd January 2025
Accepted: 27th February 2025
Published: 1st July 2025

Keywords:

Five-yearly Tax;
Text recognition;
TrOCR;
Vehicle plate

The high growth in the number of motorized vehicles in Indonesia has given rise to special attention in managing traffic administration, especially in relation to vehicle taxes. To present innovative solutions in vehicle tax administration, this research was conducted to detect the five-year tax status of motor vehicles in Indonesia using the Transformer Optical Character Recognition (TrOCR) model. The aim of this research is to evaluate the performance of the TrOCR model in recognizing text on motor vehicle number plates in Indonesia and classifying number plates that have and have not paid tax. The data used is primary data in the form of images of motor vehicle number plates taken around the Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, using a purposive sampling with constraints on the representation of each class. Although the data collection was limited to this location, Indonesian vehicle plates follow a standardized format, with regional differences primarily in the prefix letters. Additionally, the university attracts students from various regions who often use vehicles registered in their home provinces. Consequently, the collected dataset reflects a diverse range of number plates, making it a reasonable representation of motor vehicle plates across Indonesia. The research results show that the TrOCR model succeeded in achieving a Character Error Rate (CER) value of 2.9% with a data configuration of 90% for training and 10% for testing, and using 8 epochs. Evaluation of model performance indicates that overall text detection is very effective in classifying the five-year tax status of motor vehicles. Although there are some prediction errors, the overall performance of the model can be considered good and is able to provide reliable information regarding the five-yearly vehicle tax status

How to cite this article:

N. D. Fazira and A. Fauzan., "TRANSFORMER-BASED OPTICAL CHARACTER RECOGNITION APPROACH FOR IDENTIFYING MOTOR VEHICLES WITH OVERDUE TAXES," *BAREKENG: J. Math. & App.*, vol. 19, no. 3, pp. 1597-1608, September, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Vehicle plates are one of the mandatory identities for motorized vehicles, both motorcycles and cars, which are officially issued and regulated by the police. In addition to functioning as an identity attached to the vehicle, vehicle plates also function as a valid permit to operate vehicles on the highway [1]. These provisions are regulated in Law Number 22 of 2009 concerning Traffic and Road Transportation, as well as Regulation of the Chief of the Indonesian National Police Number 5 of 2012 concerning Registration and Identification of Motor Vehicles [2]. In Indonesia, vehicle plates contain identification numbers that provide important information about the vehicle, such as the type of vehicle, registration area, and year of registration [3]. The identity of the vehicle owner and related information can be identified through the vehicle plate, especially in special situations such as accidents or crimes [4], [5].

Vehicle plates play an important role in traffic administration and the vehicle taxation system. In Indonesia, vehicle owners are required to pay vehicle tax every year. Motorized Vehicle Tax is a tax on the ownership of motorized vehicles. There are two types of taxes that must be paid by vehicle owners: (1) annual tax and (2) five-yearly tax. Annual tax is a tax that is paid every year according to the due date. The five-yearly tax is a tax payment that is accompanied by the replacement of the Vehicle Registration Certificate and vehicle number plates. Funds from motorized vehicle tax managed by the provincial government are used to provide and maintain road facilities, and the construction of other public facilities that are under the authority of the regional government [6]. There are several consequences if one does not pay the motorized vehicle tax: (1) the vehicle owner will be subject to a fine, the amount of which depends on the length of the delay in paying the tax; (2) the vehicle owner can be subject to criminal sanctions in the form of imprisonment; and (3) the vehicle registration number can be removed from the system [7].

Currently, many vehicle owners do not comply with their tax payment obligations, including the five-yearly tax. Therefore, identification and verification of vehicle plates is an important effort to find out whether a vehicle has paid its five-year tax or not. Automatic vehicle plate recognition is a very important concept considering the increasing number of cars, motorbikes, and other vehicles [8]. Before the existence of sophisticated technology, the recognition of vehicle plates with overdue taxes was done manually. When the vehicle is on the highway, police officers can visually inspect the vehicle plate. They check whether the vehicle plate has fulfilled its tax obligations by checking the information on the plate or referring to the list of vehicles that have been registered as taxpayers. In addition, there are checkpoints or traffic barriers that are used to check vehicle taxes. Vehicles that pass through this checkpoint will have their vehicle plates checked to ensure compliance with tax payments.

Previous research relevant to vehicle plate detection includes a study on vehicle license plate character pattern recognition using the Momentum Backpropagation Neural Network Algorithm. The method involves preprocessing the license plate images by first converting them into binary images [9]. However, this study only focused on vehicle plate number recognition without considering the vehicle's five-yearly tax status. One widely used method for handwriting recognition is Optical Character Recognition (OCR). OCR involves the automated conversion of handwritten or printed text into machine-readable text. Some studies related to the implementation of the OCR method are as follows. [10] discussed the application of neural network-based Optical Character Recognition (OCR) to recognize handwritten manuscripts from the medieval period, particularly Carolingian Minuscule script. [11] examined the potential of Artificial Intelligence (AI) and Machine Learning (ML) technologies in OCR implementation for historical documents, focusing on early modern Spanish manuscripts. The study evaluated four primary approaches—Convolutional Neural Networks (CNN), Vision Transformers (ViT), SeqCLR (Contrastive Learning), and TrOCR (Transformer-based OCR), achieving up to 97.5% character-level accuracy. Additionally, [12] developed a CNN-based OCR system optimized to recognize individual characters in handwritten manuscripts using the Grantha script. On another note, [13] explored recent advancements in the digitization and restoration of ancient manuscripts and inscriptions, as well as the technical review studied by [14].

In addition to traditional OCR, an alternative approach to text recognition is Transformer Optical Character Recognition (TrOCR). Unlike OCR, which relies on template matching or convolutional networks, TrOCR employs a Transformer-based architecture with an encoder-decoder framework. This enables end-to-end text recognition without the need for complex preprocessing. TrOCR excels at capturing contextual relationships between characters through its attention mechanism, resulting in high accuracy even for noisy documents or complex handwriting. A study by Li et al. [15] demonstrated that TrOCR outperformed traditional OCR models across various text recognition tasks, including handwritten and printed text.

Furthermore, other research reported that TrOCR achieved word prediction accuracy rates as high as 96.4% under diverse image conditions, surpassing models such as Tesseract and CRNN **[16]**. Research by Ströbel et al. **[17]** revealed that TrOCR is capable of adapting to multilingual historical manuscript documents, even with limited training data, and consistently outperforms other Handwritten Text Recognition (HTR) systems. With its adaptability to multiple languages and flexibility in handling stylized text, TrOCR has proven to be a superior choice for complex text-based OCR tasks, including handwriting recognition and the digitization of historical documents **[18]**.

Based on the aforementioned discussion, this study aims to identify the five-year vehicle tax status in Indonesia by utilizing the Transformer Optical Character Recognition (TrOCR) method. The processing of vehicle license plate images is conducted using image processing techniques, including preprocessing steps such as grayscale conversion, character segmentation, and noise reduction, to ensure optimal recognition by the computer system. The processed data is then analyzed using the TrOCR model, known for its high precision in recognizing text patterns, even in handwritten or suboptimal image conditions. Furthermore, this study seeks to classify license plates based on their tax payment status, categorized into two main groups: vehicles with up-to-date five-year tax payments and those with overdue payments. This approach is expected to contribute significantly to improving the efficiency of vehicle tax administration in Indonesia, including enhancing oversight and law enforcement processes.

This study has the potential to enhance the efficiency of vehicle tax administration in Indonesia through the implementation of the TrOCR model, which can directly detect the five-year tax status from vehicle license plates. This automation enables faster and more accurate tax compliance inspections, supports law enforcement by facilitating real-time identification of vehicles with overdue taxes, and provides evidence-based data to inform more effective public policy development in improving vehicle tax compliance.

2. RESEARCH METHODS

The population of this study consists of all images of motor vehicle license plates within the vicinity of the Universitas Islam Indonesia. The research sample was obtained using purposive sampling, ensuring representation from each classification category. This study utilized vehicle license plates in various conditions, ranging from well-preserved to faded or damaged, and included a wide range of regional codes from across Indonesia. This approach ensures that the TrOCR model can accurately detect vehicle tax status under diverse physical conditions and geographical contexts.

2.1 Research Methodology

This study aims to develop a system that can detect the five-yearly tax status of motorized vehicles using the TrOCR method. The steps are presented in **Figure 1**. Based on **Figure 1**, the research process begins with data collection, forming the foundation for subsequent steps. The research was fully conducted in Google Colab using the Python programming language. A crucial aspect of the methodology is the splitting of the dataset into training and testing subsets, conducted across five distinct scenarios. These scenarios range from a 70%:30% to a 90%:10% split, ensuring a thorough evaluation of the model under varying data proportions. To address potential data imbalance and improve the model's generalizability, resampling techniques are applied, reinforcing the validity and robustness of the analysis **[19]**, **[20]**. The next stage involves the initialization and training of the TrOCR model using the designated training dataset. Model performance is rigorously evaluated in the fourth step by comparing Character Error Rate (CER) values, ensuring that the model achieves optimal accuracy for text recognition. Once trained, the TrOCR model is employed for license plate detection, a critical step in automating the identification process.

The initialization and training of the TrOCR model is based on a Vision Transformer (ViT) encoder and a BERT-style decoder. The model was initialized using the pre-trained "microsoft/trocr-base-stage1" checkpoint, leveraging transfer learning to enhance performance. Training was conducted using a batch size of 8 per device, a learning rate of 5e-5, and a total of 8 epochs, selected based on experimental results that indicated minimal performance improvement beyond this point while increasing the risk of overfitting.

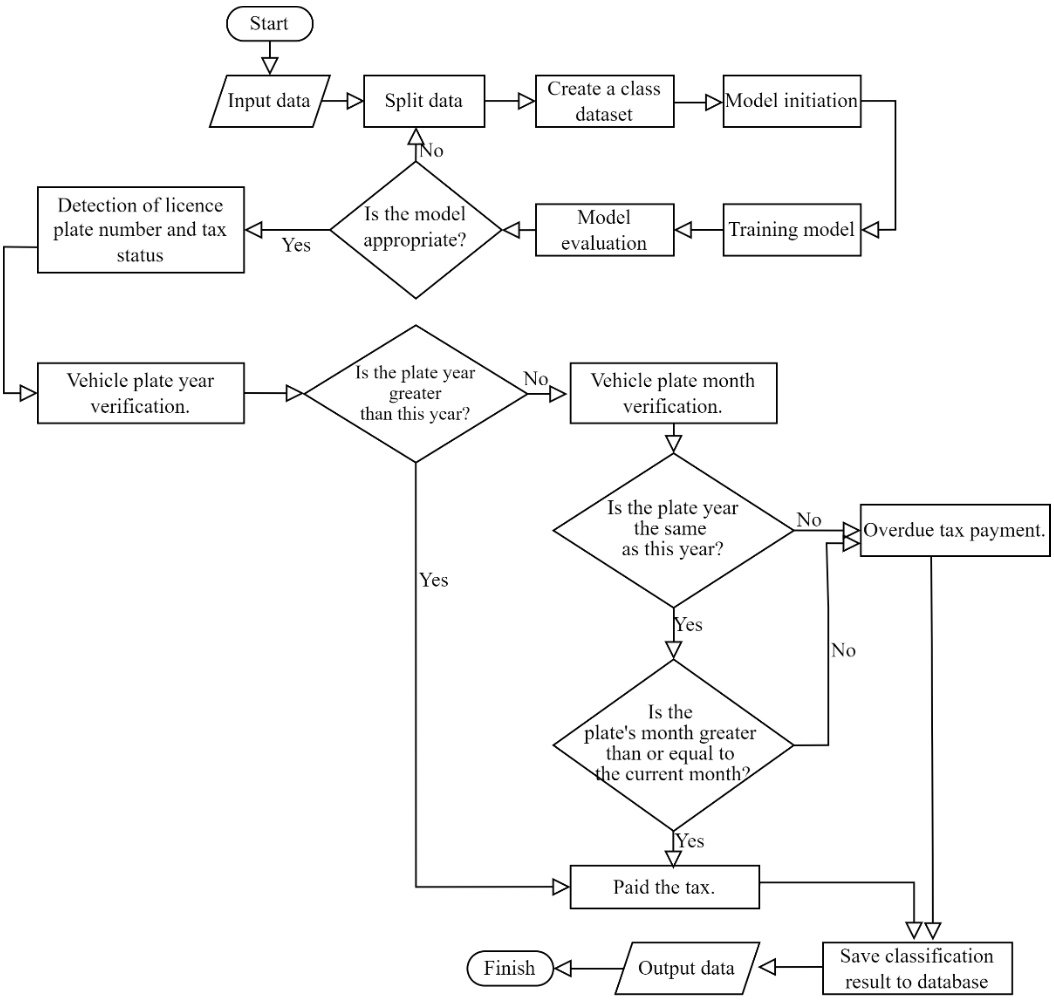

Figure 1. **Research Methodology**

To monitor training progress, the Character Error Rate (CER) and validation loss were tracked at each evaluation step. Additionally, early stopping criteria were considered to prevent overfitting, ensuring the model's stability. Once trained, the TrOCR model was employed for license plate text recognition, automating the identification process by extracting and classifying vehicle tax status based on the detected text.

Tax status classification is performed based on clearly defined criteria to ensure accurate categorization. First, if the year on the license plate matches the current year and the month is greater than or equal to the current month, the status is classified as payment completed. Similarly, if the year on the license plate is later than the current year, it is also considered as payment completed. Conversely, if the year matches the current year but the month is earlier than the current month, the status is classified as overdue. Finally, if the year on the license plate is earlier than the current year, the tax status is also classified as overdue. Upon determining the tax status, the classification results are systematically recorded in a data frame, facilitating further analysis and validation. This structured and comprehensive process not only ensures the reliability of the TrOCR model but also highlights its potential for automated vehicle tax status classification in practical applications.

2.2 Motorized Vehicle Registration Plate

Motorized Vehicle Registration Plate (MVRP) or license plate is an aluminum plate that has been registered at the One-stop Administration Services Office. Vehicle license plates in Indonesia usually begin with a letter indicating the vehicle's registration area, followed by a sequential number and alphabetical letter, for example R 3368 ZZ. Information regarding the vehicle's tax status can also be seen on the license plate, which indicates whether the tax payment is still valid or has expired. In Indonesia, vehicle license plates often have a numeric code at the end that indicates the month and year the vehicle tax is due. For example, the numeric code "07.28" indicates that the license is valid until July 2028. If the vehicle passes this due date, it will be categorized as overdue [21]. According to Law Number 28 of 2007, vehicle owners who do not pay taxes in accordance with the provisions of tax laws and regulations may be subject to administrative sanctions and/or criminal sanctions. Administrative sanctions include fines, interest, and increased rates. Criminal sanctions can be in the form of imprisonment and/or fines.

2.3 Transformer Optical Character Recognition

Deep learning is one of the techniques in machine learning that utilizes various layers of non-linear information processing for feature extraction, pattern recognition, and classification [22]. In text recognition applications, the use of deep learning opens up great opportunities to improve accuracy and efficiency in the text recognition process. This, in turn, supports the development of more sophisticated and adaptive text recognition systems. Although text recognition techniques are more focused on text processing than image processing in general, deep learning principles still play an important role in creating reliable and efficient optical character recognition (OCR) solutions [23].

TrOCR is an end-to-end text recognition system that accepts an image segmented into text lines and then performs text recognition. The system does not use convolutional layers in its architecture, but instead leverages a transformer architecture. TrOCR consists of three main steps. First, the image is resized to a fixed resolution and split into a sequence of image chunks. These chunks are used as input to a pre-trained Vision Transformer, which encodes the sequence. The encoded sequence is then decoded by a pre-trained transformer language model, generating wordpiece tokens based on the previously generated image and context. TrOCR is trained initially using synthetic data and then refined with a human-generated labeled dataset. Different combinations of encoder and decoder architectures are used to create versions of different sizes. TrOCRSMALL uses DeiTSMALL as the encoder and MiniLM as the decoder, with a total of 62 million parameters, compared to 334 million and 558 million parameters for TrOCRBASE and TrOCRLARGE, respectively [24]. The TrOCR model architecture is built on the Transformer architecture, which consists of two main components: Image Transformer for extracting visual features from text images and Text Transformer for language modeling. Figure 2 shows the schematic of the TrOCR architecture.

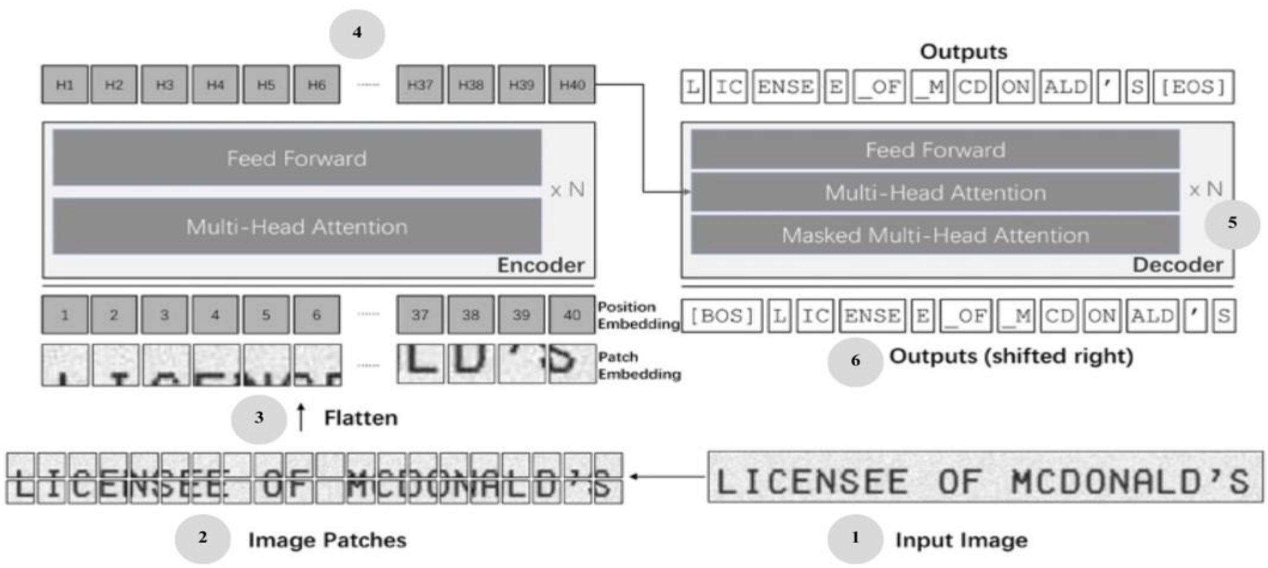

Figure 2. TrOCR Architecture [15]

The text recognition process using TrOCR, as shown in Figure 2, can be explained as follows: (1) The process begins with an image input, (2) the image is then resized to 384 × 384 pixels and divided into 576 small segments. (3) TrOCR processes the image through an encoder using the Vision Transformer (ViT) model. (4) The output of the encoder is a high-level representation of the image pieces. (5) The output of the encoder is then processed by the decoder, where BEiT translates the pieces into high-level representations, and RoBERTa gradually decomposes the representation into text characters. (6) The final output reflects the text recognition results performed by TrOCR.

2.4 Encoder and Decoder

At its basis, the encoder takes an input image $x_{img} \in \mathbb{R}^{3\times H_0\times W_0}$, and translates it into a sequence of tokens with a dimension of $H \times W$ where H represents the height of the image and W represents its width, which the transformer model can process. The encoder cannot process images that do not correspond to the token input size; the image is first resized such that its width W and height H are divisible by the patch size P. The resized image is then decomposed into a batch of $N = HW/P^2$ square patches of fixed size (P, P). These patches are then flattened into vectors and linearly projected into D-dimensional embeddings, also known as patch embeddings. Here, D represents the hidden size of the Transformer and remains consistent across all its layers, ensuring that each patch is mapped into a uniform feature space for subsequent processing **[15]**, **[25]** .

So, the image is divided into a number of square patches of size (P, P). These patches are then converted into vectors that can be linearly projected into a vector of D-dimension, which is known as the embedding patch **[15]**, **[25]**.

The decoder in the Transformer architecture plays a central role in generating meaningful output from a given input. In the context of TrOCR, the original decoder of the Transformer model is used, which consists of a stack of identical layers with an encoder-like structure. One of its key features is the encoder-decoder attention module, which allows the decoder to pay differential attention to the encoder output **[26]**. By using the keys and values from the encoder output, as well as the queries from the decoder input, this module enables deep contextual understanding. In addition, the decoder utilizes a masked self-attention mechanism to prevent the accumulation of redundant information during training. The projection of the decoder hidden states is performed through linear layers, transforming them from the model dimension to the vocabulary dimension, V. The probabilities over the vocabulary are calculated using the softmax function, which provides the required probability distribution. The beam search method is used to obtain the final output by considering multiple possibilities at once, improving the accuracy and diversity of the results. The equation related to the masked self-attention process in the decoder is described in **Equation (1)**.

$$\sigma(h_{ij}) = \frac{e^{h_{ij}}}{\sum_{k=1}^{V} e^{h_{ik}}} \text{ with } j = 1,2,\dots,V \tag{1}$$

$$h_i = Proj\big(Emb(Token_i)\big) \tag{2}$$

$\sigma(h_{ij})$ is the softmax value of h_i and is an output of the decoder's linear layer. The decoder initialization process uses the RoBERTa and MiniLM models. RoBERTa, as a modification of BERT, evaluates the impact of key hyperparameters and training data size, while MiniLM is a high-performance compressed model trained through the distillation of the self-attention module from the last layer of the Transformer on a large model (teacher). Although the decoder structure does not fully match these models, initialization is done by manually setting suitable parameters, while unsuitable parameters are randomly initialized **[25]**. With these elements, the decoder in the Transformer architecture is able to produce accurate and diverse outputs for various tasks, including text recognition on TrOCR.

2.5 Character Error Rate

Character Error Rate (CER) is an evaluation metric used to measure errors in character recognition (letters, numbers, symbols), often in the context of optical character recognition (OCR) systems **[27]**, **[28]**. In OCR, CER is used to evaluate the accuracy of transcribed text by comparing it to a reference text. The CER value can vary between 0 and 100%. The lower the CER percentage, the more accurate the encryption result **[29]**, **[30]**. **Figure 3** is a description of the types of errors in CER.

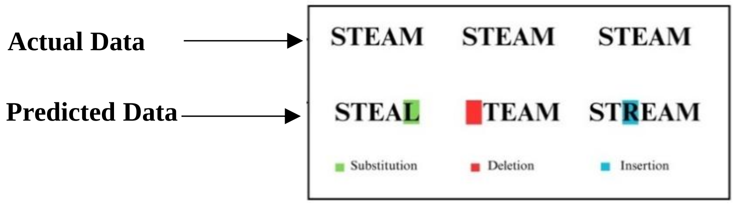

Figure 3. Types of Errors Measured by the CER [31]

According to **Figure 3**, the Character Error Rate (CER) can be classified into three types of errors: substitution (indicated in green), deletion (indicated in red), and insertion (indicated in blue). Substitution refers to the replacement of one letter with a different one. Deletion occurs when a letter is not detected by the OCR, leading to a reduction in the total number of letters in a word. Insertion, on the other hand, arises when a letter is erroneously added to a word, even though it was not originally present **[31]**. The formula for calculating the CER value is presented in **Equation (3) [31]**.

$$CER = \frac{S + D + I}{N} \times 100\% \tag{3}$$

S : number of incorrect characters, *D*: number of deleted characters, *I*: number of inserted characters, and *N*: total number of referenced characters.

3. RESULTS AND DISCUSSION

This study evaluated the effectiveness of the Transformer Optical Character Recognition (TrOCR) model in identifying the five-year tax status of motor vehicles based on license plate recognition. The results indicate that the TrOCR model successfully achieved a Character Error Rate (CER) of 2.9% with a dataset split of 90% training and 10% testing, using 8 training epochs. This demonstrates the model's ability to recognize and classify vehicle tax status with high accuracy. Additionally, the method proved effective in handling diverse image conditions, including faded or damaged plates. The following sections provide a detailed breakdown of the data processing steps, model evaluation, and classification performance.

3.1 Image Cropping and Data Split

Image cropping is done to direct the focus on the vehicle plate surface, making it easier for the computer to process data and detect text. This cropping process makes it easier for the machine to recognize text, reduces the level of prediction errors, and speeds up processing time **[32]**. This study stores vehicle plate image data in one folder and names it with a number for identification purposes. Each image is named by number to facilitate the creation of text identities in CSV files. **Figure 4** shows part of the license plate dataset.

Filename	Identity
1.jpg	AB1895KA0918
2.jpg	BE6526LC1027
3.jpg	AB1846HS1024
4.jpg	BH2430RD1027

(a) **(b)**

Figure 4. (a) Collection of License Plate Images, (b) Identity of Each Plate Image

The dataset will be systematically divided into training and testing data, with the allocation of each determined based on predefined proportions. Notably, the proportions of training and testing data are carefully simulated, ranging from 70%:30% to 90%:10%, ensuring a comprehensive evaluation of the model's performance under varying data splits. This step is critical to balance the trade-off between model training and validation accuracy, providing robust insights into its effectiveness.

3.2 Image Labelling

The image labeling process begins with creating a 'PlateDataset' class that includes the labeling of vehicle plate images. Once the images are labeled, the text in the images is then tokenized, which means the text is converted into a sequence of numeric tokens for easy processing by the model. The steps in the tokenization process include: (1) text tokenization to convert each character into a numeric token, (2) text padding to ensure consistency of input length, (3) handling padding tokens by replacing their values to exclude them from the loss calculation during model training, and (4) dictionary encoding representation of the preprocessing results, which includes the representation of images and labels in tensor form for use in model training. Image labeling is a crucial step because, without labels, the model cannot learn effectively, cannot calculate errors, and is unable to provide accurate predictions. This will negatively impact the quality of the model and limit its ability to be used in real applications [33], [34]

3.3 Model Construction

The TrOCR model formation process consists of two main stages, namely: (1) initialization and (2) model training. The initialization process of the TrOCR model using the Transformer architecture involves the necessary initial configuration and settings to ensure compatibility with the model used. A model object from the VisionEncoderDecoderModel class is created using the pre-trained model "microsoft/trocr-base-stage1 [15]. Several parameters, such as the initial decoder token, token padding, vocabulary size, and beam search parameters, were adjusted according to the research needs. The model training process begins by setting training arguments for the model. The arguments set includes batch sizes for training and evaluation, device memory settings, and model storage frequency. This study assessed the effect of training and testing data split ratios, as well as the number of epochs used.

The TrOCR model formation process consists of two main stages: (1) initialization and (2) model training. The initialization process of the TrOCR model follows the Transformer-based VisionEncoderDecoderModel architecture. A model object was created using the pre-trained "microsoft/trocr-base-stage1" checkpoint, which includes a Base-sized decoder with 12 layers**,** a hidden size of 1024, and 16 attention heads. The decoder also utilizes a feed-forward network dimension of 4096 and a dropout rate of 0.1, ensuring efficient text generation. Several parameters were adjusted to optimize performance. The beam search size was set to 4**,** with a no-repeat n-gram size of 3 and a length penalty of 2.0**,** balancing accuracy and diversity in text generation. Additionally, modifications were made to the initial decoder token, token padding, and vocabulary size (50265 tokens) to align with the dataset's characteristics.

The model training process involved fine-tuning on a custom license plate dataset, leveraging transfer learning from the pre-trained TrOCR model. The training setup used a batch size of 8 per device, a learning rate of 5e-5, and 8 training epochs, selected based on experiments evaluating the trade-off between model convergence and overfitting risks. Training and evaluation followed multiple data split scenarios to assess the impact of dataset proportions. The effectiveness of fine-tuning was monitored through CER, which tracks the model's performance across training iterations. The results showed a progressive decrease in CER, confirming that fine-tuning improved text recognition, especially for license plates. This demonstrates the model's robustness for real-world vehicle tax classification applications. This study attempts to evaluate the effect of various split ratios of the training and testing data on the model training process. The results of each ratio can be seen in **Table 1**. Based on **Table 1**, the data split ratio of 90% for training and 10% for testing achieved better performance, so this ratio was chosen for use in this study.

Table 1. Train-Test Ratio Comparison

Split Ratio (Train : Test)	CER		Loss Validation	Time (Second)
	Train	Test		
70% : 30%	1.9%	6%	158	448
75% : 25%	1.9%	8.8%	169	383
80% : 20%	1.5%	8.2%	180	394
85% : 15%	0.7%	10%	192	417
90% : 10%	0	2.9%	230	570

The second model parameter studied is the number of epochs. In this study, researchers compared the CER values of models trained with 5, 8, and 10 epochs. **Table 2** shows the results of the comparison of CER values based on the number of epochs used in training, with a data split ratio of 90% for training and 10% for testing.

Table 2. Result Epoch

	Epoch 5	Epoch 8	Epoch 10
CER	7.8%	2.9%	5.2%
Training Loss	0.350	0.20	0.660
Time	295	570	617
Validation Loss	0.485	0.369	0.509

The results in **Table 2** show that the use of 8 epochs produces the lowest CER value, which is 2.9%, and the lowest training loss value, which is 0.204. Therefore, the number of epochs to be used in this study is 8. Interestingly, increasing the epochs from 5 to 8 improved accuracy, but further increasing to 10 epochs resulted in a higher CER and loss. Theoretically, a larger number of epochs allows the model to better learn patterns from the data, leading to improved accuracy. However, after 8 epochs, the model likely begins to overfit the training data, memorizing its patterns instead of generalizing well to new data. This explains why epoch 8 performed better than epoch 10 in this study. Based on the model obtained, the average CER for training data and testing data based on **Equation (3)** was 0 for CER training data and 0.029 for testing data. Recognition of the five-yearly tax due date (last 4 digits) performed on training data resulted in no errors. On the other hand, the results of the five-yearly tax recognition on the test data show that there are 2 incorrectly predicted images, which are: (1). The 17^{th} data which should have a value of 1026, but is predicted as 1023, (2) the 22nd data which should have a value of 1124, but is predicted as 0124. Based on this error, the CER value for the recognition of the five-year tax due dates for the test data can be calculated as 0.021. The incorrect prediction could be due to blurred characters, variations in license plate fonts, and lighting conditions affecting text visibility.

3.4 Out-of-Sample Prediction

The TrOCR model was tested on 5 out-of-sample images. Prediction includes recognition of the license plate number and the five-yearly tax status. The five-yearly tax status is classified based on the following rules: (1) If the year on the license plate is the same as the current year and the month on the license plate is greater than or equal to the current month, then it is classified as having paid tax. (2) If the year on the license plate is greater than the current year, then it is classified as having paid tax. (3) If the year on the license plate is the same as the current year and the month on the license plate is less than the current month, then it is classified as overdue. (4) If the year on the license plate is less than the current year, then it is classified as overdue. By applying this classification, the results of license plate number recognition and tax status classification on out-of-sample data are presented in **Table 3**.

Table 3. Results of Prediction

No	Image	Prediction	Number of Errors	Tax Payment Status Prediction	Notes
1		E3373ZG1026	0	The registered plate number has paid the five-yearly tax	-
2		BA4882IB0926	0	The registered plate number has paid the five-yearly tax	-
3		Z6341K0827	0	The registered plate number has paid the five-yearly tax	-
4		B6040NTN1023	1	The registered plate number has not paid the five-yearly tax	Error in recognizing the character T, which was instead read as N by the computer

No	Image	Prediction	Number of Errors	Tax Payment Status Prediction	Notes
5		AE4808IH0227	0	The registered plate number has paid the five-yearly tax	-

Analysis of the prediction results in **Table 3** concludes that the TrOCR method applied in this study showed very satisfactory performance. The majority of the data was detected correctly, and there were only a few errors in character recognition. These results indicate that TrOCR is able to handle various forms of text contained in the image. The error rate in the detection process may be caused by the similarity between characters, which results in inaccurate predictions. However, this error is very minimal and does not significantly affect the final results of the five-year vehicle tax status detection. This study shows that TrOCR can be relied on for data processing that requires high accuracy, such as tax status detection. With almost perfect accuracy, the application of this model should improve the efficiency of the vehicle tax detection process.

4. CONCLUSIONS

This study applies the Transformer Optical Character Recognition (TrOCR) method to detect and extract text from vehicle license plate images in Indonesia. A key finding of the research highlights that the model achieved its optimal performance by utilizing a 90% training and 10% testing data split, combined with 8 training epochs, resulting in a Character Error Rate (CER) of just 2.9%**.** This optimal configuration demonstrates the model's capability to balance training efficiency and prediction accuracy effectively. The recognized text is subsequently employed to identify vehicle license plates and classify their five-yearly tax status. Despite minor prediction errors, the model exhibits strong overall performance in providing accurate and actionable information regarding vehicle tax status. These results emphasize the potential of TrOCR for OCR-based information systems, particularly in the domain of vehicle tax management and monitoring in Indonesia. Nevertheless, further research is essential to enhance prediction accuracy and broaden the dataset scope, ensuring reliability under diverse conditions and data variations. Despite minor prediction errors, the model demonstrates strong performance in accurately identifying vehicle tax status. However, its effectiveness is influenced by image quality and text clarity, as variations in lighting, shadows, and faded or obstructed characters can impact recognition accuracy. Future research should improve dataset diversity by incorporating different lighting conditions, worn-out plates, and various angles. Further optimizations in preprocessing and model architecture could also enhance performance in real-world scenarios.

ACKNOWLEDGMENT

The author would like to express gratitude to the Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia, for their invaluable assistance in facilitating the smooth conduct of this research.

REFERENCES

[1] Kumparan, "ARTI PLAT NOMOR DAN SERBA-SERBI IDENTITAS KENDARAAN." [Online]. Available: https://kumparan.com/info-otomotif/arti-plat-nomor-dan-serba-serbi-identitas-kendaraan-1xBytPuVTLT/4

[2] M. A. Taufik, "TINJAUAN HUKUM ISLAM TERHADAP JUAL BELI PLAT NOMOR KENDARAAN BERMOTOR DI YOGYAKARTA," *Az Zarqa'*, vol. 7, no. 2, pp. 237–251, 2015.

[3] Baladewa, "ARTI DAN JENIS PLAT NOMOR KENDARAAN DI INDONESIA." [Online]. Available: https://www.bhinneka.com/blog/arti-dan-jenis-plat-nomor-kendaraan-di-indonesia/

[4] N. Andari, "DASAR HUKUM DAN PERATURAN PLAT NOMOR KENDARAAN," Carmudi Indonesia. [Online]. Available: https://www.carmudi.co.id/journal/dasar-hukum-dan-peraturan-plat-nomor-kendaraan/

[5] E. Wattimury, H. Z. Wadjo, and E. Ubwarin, "PENGGUNAAN PELAT TANDA NOMOR KENDARAAN BERMOTOR YANG DIPALSUKAN DI KOTA AMBON," *LUTUR Law Journal*, vol. 2020, no. 1, 2020.

[6] Kantor Pelayanan Pajak Daerah, "PAJAK 5 TAHUNAN." [Online]. Available: https://samsatsleman.jogjaprov.go.id/index.php/layanan/pajak-5-tahunan

[7] BAPENDA JABAR, "CARA MENGHITUNG PAJAK PROGRESIF KENDARAAN BERMOTOR." [Online]. Available: https://bapenda.jabarprov.go.id/2017/03/14/cara-menghitung-pajak-progresif-kendaraan-bermotor/

[8] Y. Galahartlambang, T. Khotiah, Z. Fanani, and A. A. Y. Solekhah, "DETEKSI PLAT NOMOR KENDARAAN OTOMATIS DENGAN CONVOLUTIONAL NEURAL NETWORK DAN OCR PADA TEMPAT PARKIR ITB AHMAD DAHLAN LAMONGAN," *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, vol. 6, no. 2, pp. 114–122, 2023, https://doi.org/10.36595/misi.v5i2.

[9] D. Avianto, "PENGENALAN POLA KARAKTER PLAT NOMOR KENDARAAN MENGGUNAKAN ALGORITMA MOMENTUM BACKPROPAGATION NEURAL NETWORK," *Jurnal Informatika*, vol. 10, no. 1, 2016.

[10] B. W. Hawk, A. Karaisl, and N. White, "MODELLING MEDIEVAL HANDS: PRACTICAL OCR FOR CAROLINE MINUSCULE," *DHQ: Digital Humanities Quarterly*, vol. 13, no. 1, 2019.

[11] A. Khan *et al*., "OCR APPROACHES FOR HUMANITIES: APPLICATIONS OF ARTIFICIAL INTELLIGENCE/MACHINE LEARNING ON TRANSCRIPTION AND TRANSLITERATION OF HISTORICAL DOCUMENTS," *Digital Studies in Language and Literature*, vol. 1, no. 1–2, pp. 85–112, Dec. 2024, https://doi.org/10.1515/dsll-2024-0013.

[12] A. Jindal and R. Ghosh, "AN OPTIMIZED CNN SYSTEM TO RECOGNIZE HANDWRITTEN CHARACTERS IN ANCIENT DOCUMENTS IN GRANTHA SCRIPT," *International Journal of Information Technology*, vol. 15, no. 4, pp. 1975–1983, Apr. 2023, https://doi.org/10.1007/s41870-023-01247-1.

[13] N. Jayanthi, S. Indu, S. Hasija, and P. Tripathi, "DIGITIZATION OF ANCIENT MANUSCRIPTS AND INSCRIPTIONS - A Review," 2017, pp. 605–612. https://doi.org/10.1007/978-981-10-5427-3_62.

[14] P. Bhatt and I. P. Patel, "OPTICAL CHARACTER RECOGNITION USING DEEP LEARNING," vol. 11, no. 1, pp. 55–66, 2018, [Online]. Available: https://www.researchgate.net/publication/326009476_Optical_Character_Recognition_using_Deep_learning_-_A_Technical_Review

[15] M. Li *et al*., "TrOCR: TRANSFORMER-BASED OPTICAL CHARACTER RECOGNITION WITH PRE-TRAINED MODELS," in *The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)*, 2023. [Online]. Available: www.aaai.org

[16] R. L. Zhang, "A COMPREHENSIVE EVALUATION OF TROCR WITH VARYING IMAGE EFFECTS," 2024.

[17] P. B. Ströbel, S. Clematide, M. Volk, and T. Hodel, "TRANSFORMER-BASED HTR FOR HISTORICAL DOCUMENTS," *arXiv preprint arXiv:2203.11008*, 2022.

[18] L. Li, "HANDWRITING RECOGNITION IN HISTORICAL DOCUMENTS WITH MULTIMODAL LLM," *arXiv preprint arXiv:2410.24034*, 2024.

[19] J. X. C. Ke, A. DhakshinaMurthy, R. B. George, and P. Branco, "THE EFFECT OF RESAMPLING TECHNIQUES ON THE PERFORMANCES OF MACHINE LEARNING CLINICAL RISK PREDICTION MODELS IN THE SETTING OF SEVERE CLASS IMBALANCE: DEVELOPMENT AND INTERNAL VALIDATION IN A RETROSPECTIVE COHORT," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 91, Nov. 2024, https://doi.org/10.1007/s44163-024-00199-0.

[20] *Resampling Methods*. Boston, MA: Birkhäuser Boston, 2006. https://doi.org/10.1007/0-8176-4444-X.

[21] S. Kehi, S. E. M. Nirahua, and H. M. Y. Tita, "PENYALAHGUNAAN PEMBAYARAN PAJAK KENDARAAN BERMOTOR PEMERINTAH BERDASARKAN PERATURAN DAERAH NOMOR 6 TAHUN 2010," *TATOHI: Jurnal Ilmu Hukum*, vol. 2, no. 9, p. 966, Aug. 2022, https://doi.org/10.47268/tatohi.v2i9.1436.

[22] L. Deng and D. Yu, "DEEP LEARNING: METHODS AND APPLICATIONS," *Foundations and Trends R in Signal Processing*, vol. 7, pp. 197–387, 2013, https://doi.org/10.1561/2000000039.

[23] A. Kamilaris and F. X. Prenafeta-Boldú, "DEEP LEARNING IN AGRICULTURE: A SURVEY," *Comput Electron Agric*, vol. 147, pp. 70–90, Aug. 2018, https://doi.org/10.1016/J.COMPAG.2018.02.016.

[24] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "MINILM: DEEP SELF-ATTENTION DISTILLATION FOR TASK-AGNOSTIC COMPRESSION OF PRE-TRAINED TRANSFORMERS," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, ancouver, Canada, Aug. 2020. [Online]. Available: http://arxiv.org/abs/2002.10957

[25] H. Touvron *et al*., "TRAINING DATA-EFFICIENT IMAGE TRANSFORMERS & DISTILLATION THROUGH ATTENTION," in *Proceedings of the 38 th International Conference on Machine Learning*, PMLR, 2021.

[26] A. Vaswani *et al*., "ATTENTION IS ALL YOU NEED," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA., 2023.

[27] C. Neudecker, K. Baierer, M. Gerber, C. Clausner, A. Antonacopoulos, and S. Pletschacher, "A SURVEY OF OCR EVALUATION TOOLS AND METRICS," in *Proceedings of the 6th International Workshop on Historical Document Imaging and Processing*, in HIP '21. New York, NY, USA: ASSOCIATION FOR COMPUTING MACHINERY, 2021, pp. 13–18. https://doi.org/10.1145/3476887.3476888.

[28] Y. Baek *et al*., "CLEval: CHARACTER-LEVEL EVALUATION FOR TEXT DETECTION AND RECOGNITION TASKS," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.06244

[29] Muslih and L. B. Handoko, "PENGUJIAN AVALANCHE EFFECT PADA KRIPTOGRAFI TEKS MENGGUNAKAN AUTOKEY CIPHER," in *2 st Proceeding STEKOM*, Stekom University, 2022.

[30] D. Das, J. Philip, M. Mathew, and C. V Jawahar, "A COST EFFICIENT APPROACH TO CORRECT OCR ERRORS IN LARGE DOCUMENT COLLECTIONS," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 655–662. https://doi.org/10.1109/ICDAR.2019.00110.

[31] K. Leung, "EVALUATE OCR OUTPUT QUALITY WITH CHARACTER ERROR RATE (CER) AND WORD ERROR RATE (WER)." [Online]. Available: https://towardsdatascience.com/evaluating-ocr-output-quality-with-character-error-rate-cer-and-word-error-rate-wer-853175297510

[32] Yuhandri, "PERBANDINGAN METODE CROPPING PADA SEBUAH CITRA UNTUK PENGEMBILAN MOTIF TERTENTU PADA KAIN SONGKET SUMATERA BARAT," *Jurnal KomtekInfo*, vol. 6, no. 1, pp. 96–105, 2019, [Online]. Available: http://lppm.upiyptk.ac.id/ojsupi/index.php/KOMTEKINFO

[33] O. Russakovsky *et al.*, "IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015, https://doi.org/10.1007/s11263-015-0816-y.

[34] Y. Lecun, Y. Bengio, and G. Hinton, "DEEP LEARNING," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, https://doi.org/10.1038/nature14539ï.