TIMOR-LESTE'S MIKROLET VEHICLE MULTI-CLASS CLASSIFICATION PERFORMANCE EVALUATION USING DEEP LEARNING MODELS

Abreu Andre Boavida

Faculty of Engineering, Universidade Nacional Timor-Lorosae, Hera, East Timor, Graduate school of Natural Science and Technology, Gifu University - Japan

ABSTRACT

This research uses a multiclass deep learning model to identify images containing different mikrolet vehicles. In addition, we have studied multiclass classification research by assessing the performance of mikroletpictures with the same color and comparing the classification performance of several proposed deep learning models. The objectives of this study are two-fold: first, to identify images containing different mikrolet vehicle lanes, and second, compare four deep learning models for evaluation of the accuracy of performance precision, recall, f1-measure, micro and macro average ROC curves areas.

KEYWORDS

Deep Learning Models, Classification, Mikrolet, Multiclass, Vehicles, ROC

1. INTRODUCTION

Transportation is the movement of people and goods from one place to another using motorized vehicles or human and animal power [1,2].

The transportation system in Timor-Leste consists of various modes of transportation, including buses, taxis, and Mikrolet vehicles. Accurate vehicle classification is crucial for effective traffic management, as it allows for a better understanding of traffic patterns and congestion. This information to improve transportation infrastructure and reduce traffic-related issues. Deep learning models can accurately classify vehicles, leading to more efficient and effective traffic management.

Deep learning, a burgeoning Artificial Intelligence (AI) technology, has significantly improved the state-of-the-art in image recognition, speech recognition, and navigation [3, 4]. The astounding methodology has also been applied to various transportation fields to enhance the management of various vehicle problems. Multiple studies have shown that in deep learning algorithms performed in the last few decades, multiple types of sensors have been utilized for vehicle detection and classification. The detection and classification of vehicles are significant for transportation monitoring and bridge management. The vehicle information is collected for vehicle monitoring and control, data analysis and visualization, and transportation system improvement [5]. For example, the vehicle location and weight are essential for the weight-limit

David C. Wyld et al. (Eds): SOEN, SIPP, PDTCA, ITE, CCSIT, NLPCL, DaKM, BIGML, AISC -2023
pp. 141-152, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131212

inspection. Besides that, accurate vehicle counts, and classification are required for intelligent transportation systems (ITS) and the performance evaluation of bridge structures [6]. In image classification, many deep learning models have been proposed by researchers [7, 8]. Therefore, it is essential and efficient for us to know the performance of these networks to provide transplant instructions with objective evaluation indexes. As the population increases, the government will start adding transportation lanes to twelve lanes according to their respective colors and numbers. Therefore, in Dili city of Timor-Leste, most people use public transportation is the mikrolet. However, there are twelve classes of mikrolet vehicles, but no real-time Navi system in Dili city, which has caused great inconvenience to everyone's travel, and it is also easy to cause traffic jams.

However, the public transportation system in Dili city relies on traditional classification methods that may not accurately classify vehicles in complex traffic situations. Officers need help to predict the path of public cars, especially in mikrolet transportation. This issue highlights the need for more advanced classification methods in the transportation system of Dili city. In this paper, we propose using deep learning models to classify mikrolet vehicles in Timor-Leste multiclass.

The aims of the study are (a) identifying images containing separate lanes of mikrolet cars for micro and macro average ROC area. (b) comparing four deep learning models for multiclass classification of mikrolet vehicles and evaluating the performance of each model. (c) to assess the performance of these models in accurately classifying cars and improving traffic management in the region.

In the following, researchers present some similar studies and emphasize the findings in section 1. Research materials and methods. 2. We provide evaluation performance results 3. In the last section, we provide conclusions and future work.

2. Material and Methods

The block diagram of the comparison model deep learning multi-classification mikrolet vehicle is shown in the figure. 1. The model consists of three main stages: data pre-processing, deep learning model for feature extraction, and classification. The data pre-processing includes Augmentation, resizing, crop, and split. In the deep learning part, four models are considered, CNN, VGG16, ResNet50, and DenseNet121. The input of the model is a mikrolet image. The final output is the classification, and evaluation performance of twelve mikrolet lines.

2.1. Dataset Mikrolet Vehicles

Figure 2. The sample dataset mikrolet used in this study contains images of twelve different mikrolet lines. The dataset was pre-processed using Augmentation, resizing, crop, and split techniques before being used for training and testing the deep learning models. The dataset was then split into training, validation, and testing sets to ensure the models were trained and evaluated on different data sets. The mikrolet dataset used in this study contains a total of 14,000 images. The images are distributed across three sets: training set, validation set, and testing set. The training set contains 12000 images, the validation set includes 350 images, and the testing set contains 1650 images. The distribution of ideas across the groups is done randomly to ensure that the models are trained and tested on diverse images.

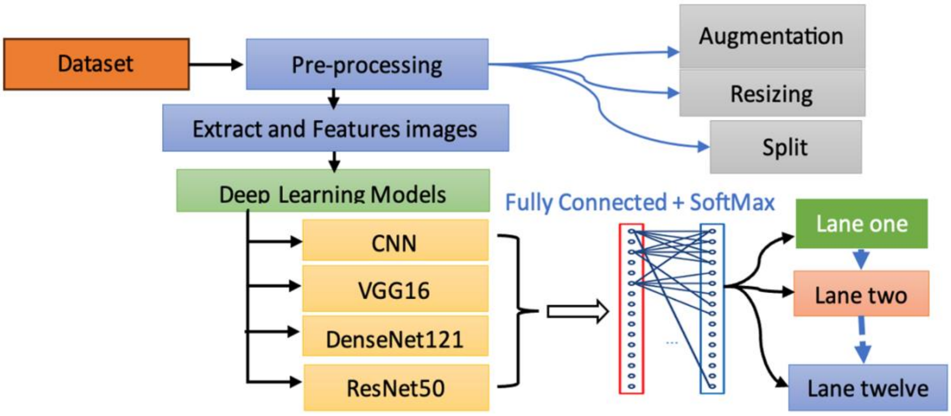

Figure 1. Block Diagram of the Comparison Model Deep Learning Multi-Classification Mikrolet Vehicle

Figure 2. Sample dataset mikrolet

2.2. Pre-Processing

The pre-processing stage of the mikrolet dataset includes several steps, such as Augmentation, resizing, and split. Augmentation involves creating new images by applying transformations such as rotation, flipping., and zooming to the existing images. Resizing involves changing the size of the images to a standard size, which in this case is 1920 x 1080 pixels. Crop consists in selecting a specific region of interest in the picture. Split involves dividing the dataset into training and testing sets.

2.2.1. Augmentation

The augmentation process for the mikrolet dataset involves applying various transformations to the original images to increase the size and diversity of the dataset. This includes techniques such as rotation, flipping, and scaling. We used the augmentation technique to create ten more images on each available [9]. In Table 1, Data augmentation is described for the parameters used for Augmentation.

Table 1. Data Augmentation

Rotation Range	10 Degree
Width shift range	0.1 Degree
Heigh shift range	0.1 Degree
Shear range	0.15 Degree
Zoom range	0.5, 1.5
Channel Shift range	150.0

2.2.2. Resizing Images

That the original images in the mikrolet dataset were of standard dimensions of 1920 x 1080 pixels. However, to reduce the computational complexity and improve the training efficiency, the images were resized to a smaller size of 256x256 pixels. This was done using standard image processing techniques such as bilinear interpolation for resizing. Show in figure 3. (a,b) resizing images mikrolet. Maintaining the aspect ratio when resizing the image is essential to avoid distortion.

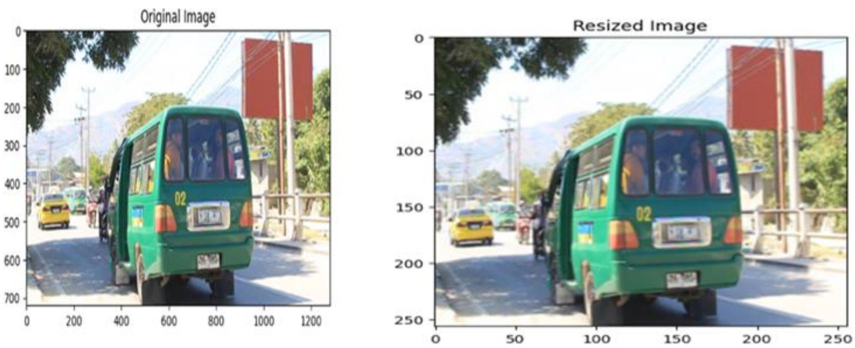

Figure 3. Figure 3. (a) Original mikrolet images. (b) Resizing images

2.2.3. Split Data

The data splitting process in this paper involves dividing the dataset into separate subsets for training with 12000 images, validation with 350 photos, and testing with 1650 images. This helps to evaluate the model's performance and make improvements if necessary.

2.3. Deep Learning Methods

Deep learning is defined as representing data in multiple and successive layers. The number of layers in deep understanding is an essential criterion for describing the depth of the network. Three main developments impact the popularity and effectiveness of deep learning.

2.3.1. Convolution Neural Network (CNN) Model

The convolutional neural network number of layers extracts local features on large-dimension data. Each layer consists of distinct nodes with learnable bias and weights; in the convolutional layer’s connection, weights are shared and called a convolutional kernel. A different number of activation functions decides all the operation results. In CNN, the previous layer's output is convolved with a learnable kernel, and weights-sharing plays a crucial role in training to reduce the number of weights. The general formula for a convolutional layer is given by equation (1-3):

$$Height = \frac{Imageheight - kernelheight + 2(padding)}{strides} + 1, \quad (1)$$

$$Width = \frac{Imagewidth - kernelwidth + 2(padding)}{strides} + 1, \quad (2)$$

$$C = \frac{W-K+2P}{S} + 1, \quad (3)$$

The convolution layer W denotes the image height and width, K represents the filter size, P is the padding, and S refers to the strides.

In Figure 4. A CNN visual representation takes the mikrolet images as input. It indicates the relationships between the layers of a CNN model to determine the class of the mikrolet images. In the convolution layer, the feature is extracted from the image with the help of various filters. The intermediate process is applied between the convolution and pooling layers so that the parts are non-linear with the Rectified Linear Unit activation function. In the pooling layer, the dimensions of these feature maps are reduced, which reduces the computational workforce in the subsequent layers and displays the features in the image more effectively. The final layer of the convolutional neural network is in the form of a classical, fully connected artificial neural network. In this layer, a fully connected structure between the artificial neurons represents the features of the image/text and the target class labels. The new mikrolet images serve as an input to the CNN. When the training is completed, CNN provides the predictive class probability.

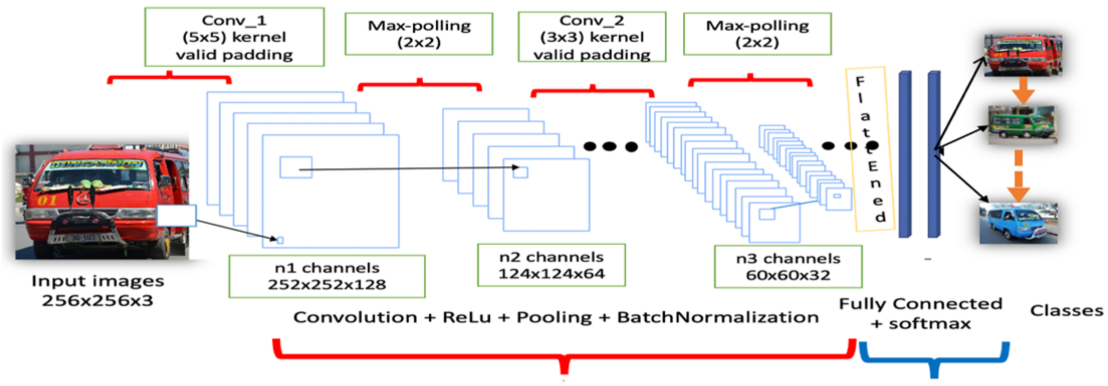

Figure 4. Architecture CNN for Mikrolet images Classification

2.3.2. VGG16 model

The VGG16 model is a convolutional neural network trained on the ImageNet dataset. This study uses an input image size of 256x256 pixels, and we need to resize the image to that size before passing it through the model. It will then process the image through a series of convolutional and pooling layers to extract features before passing it through several fully connected layers to make predictions. More details are shown in Figure 5. The main characteristics of VGG nets [32] include using 3×3 convolution layers, which improved network performances and made the network deep. 3×3 receptive filters were used throughout the net with strides of 1. The input has a shape of $256 \times 256 \times 3$. In the present work, we try more on the VGG16 model using our datasets, employing the Adam optimizer and Rectified Linear Unit (ReLU) activation function.

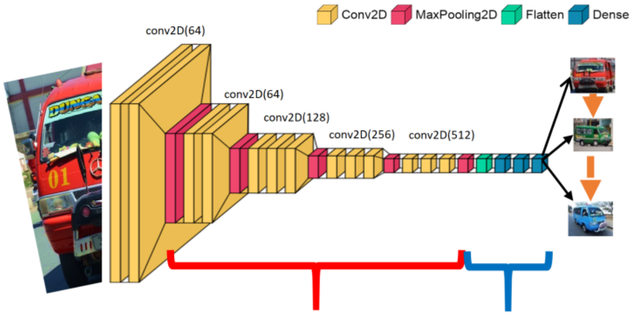

Fig. 5 Architecture VGG16 Model

2.3.3. Densenet121 Model

The fundamental blocks in the DenseNet architecture are connected densely to each other, leading to low computational requirements since the number of trainable parameters decreased heavily. Due to its narrow architecture, the DenseNet architectures add small sets of feature maps. We used the DenseNet-121 variant with the Adam optimizer and ReLU activation function for in-depth feature extraction. There are three kinds of blocks in the DenseNet implementation:

1. Convolution block, which is a basic block of a dense block. The convolution block is like the identity block in ResNet.
2. Dense block, in which convolution blocks are concatenated and densely connected. The thick block is the main component in DenseNet.
3. Transition layer, which connects two contiguous dense blocks.

Since feature map sizes are the same within the dense block, the transition layer reduces the dimensions of the feature map. The technique of bottleneck design is adopted in all the blocks. The architecture of DenseNet121 is shown in Figure 6.

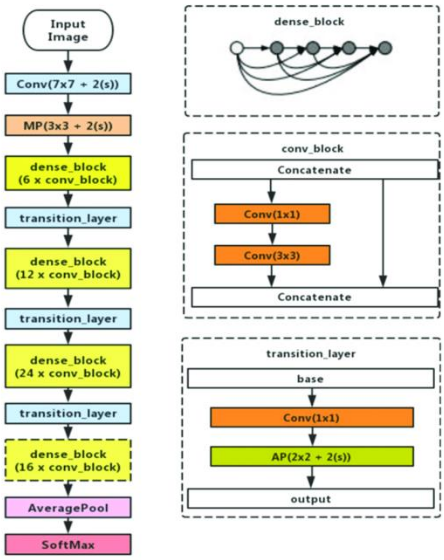

Figure 6. (**Left**) DenseNet121 architectures. (**Right**) Dense block, convolution block and transition layer

2.3.4. ResNet50 Model

ResNet50 is another deep representative network. However, deep networks are hard to train because of the notorious vanishing gradient problem as the gradient is backpropagated to earlier layers; repeated multiplication may make the angle infinitely small. By introducing a skip connection (or shortcut connection) to fit the input from the previous layer to the next layer without any input modification, ResNet50 can have an intense network of up to 152 layers. The architecture of ResNet50 is shown in Figure 3. There are two kinds of shortcut modules in the implementation of ResNet50. The first is an identity block with no convolution layer at the shortcut. In this case, the input has the exact dimensions as the output. The other is the convolution block, which has a convolution layer at the shortcut. In this case, the input dimensions are smaller than the output dimensions. In both leagues, 1 × 1 convolution layers are added to the start and end of the network. This technique, called bottleneck design, reduces the number of parameters while not degrading the network's performance so much.

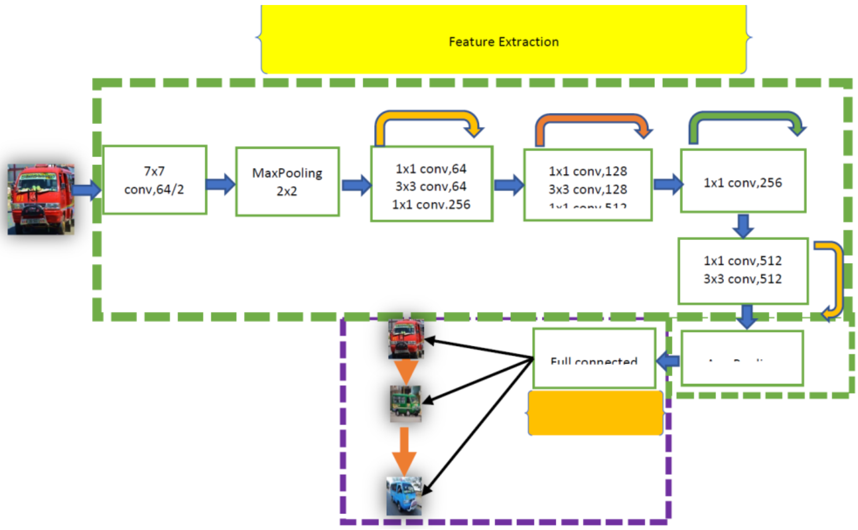

Fig. 7 Architecture of ResNet50 Model

2.4. Mulit-Class Classification Mikrolet

Multiclass classification of mikrolet vehicles involves training a machine learning model to classify images of mikrolet into multiple categories. This can be done using convolutional neural networks designed to recognize image patterns. The model is trained on a labeled image dataset and then tested on a separate set of ideas to evaluate its accuracy. This can be useful for traffic monitoring and public transportation planning applications.

3. EVALUATION PERFORMANCE AND DISCUSSION

The evaluation performance of Mikrolet vehicles can be measured using various metrics such as accuracy, precision, recall, and F1 score. These metrics can be obtained by comparing the predicted labels of the Mikrolet vehicles with their actual titles. The confusion matrix can also be used to evaluate the performance of the Mikrolet vehicle classification model.

3.1. Performance Metrics

The comparison of measurements obtained from a confusion matrix to classification achievements in similar studies. The accuracy, precision, and F1 measurement values can be obtained from the confusion matrix. A simple confusion matrix for a multiclass classifier is given in Table 2

Table 2. Multiclass classification problem confusion matrix

		Predicted class			
		C_1	C_2	…….	C_N
Actual Class	C_1	$C_{1,1}$	FP	. . .. . .	$C_{1,N}$
	C_2	FN	TP	. . . . . .	FN
	. . . .. .	………	……….	……..	……
	C_N	$C_{N,1}$	FP	…… …	$C_{N,N}$

3.2. Multiclass Confusion Matrix and Metrics

Therefore, the characterization of TP, TN, FP, and FN instances is not applicable in this case. Instead, it is feasible to perform an analysis focusing on a specific class based on the characterization provided in the example of Table 2. A set of metrics for each class can be defined based on this approach. Then, based on the proper combination of these metrics, it is feasible to provide metrics for the entire confusion matrix. Table 3 provides an overview of the metrics defined for a multiclass confusion matrix and accuracy, recall, precision, and F1-score. As can be seen equation 4- 7, there are two main approaches to defining performance metrics: "micro" and "macro."

Table 3. Performance metrics for a multiclass confusion matrix

Metric	Formula
Accuracy	$Acc(A_{reduced}) = \frac{\sum_{i=1}^{N} TP(C_1)}{\sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}}$ (4)
Accuracy of Class C_i	$TPR(C_1) = \frac{TP(C_i)}{TP(C_i)+FN(C_i)}$ (5)
The precision of Class C_i	$PPV(C_1) = \frac{TP(C_i)}{TP(C_i)+FP(C_i)}$ (6)
$F_1 - Score$ of C_i	$F_1(C_1) = 2x\frac{TPR(C_i)xPPV(C_i)}{TPR(C_i)+PPV(C_i)}$ (7)

3.2.1. Confusion Matrix

The confusion matrix represents the number of confusions by the classifier in predicting each category of kidney images. Also, it is understood from the figure. 8 (a-d), most lanes one, lane two, lane three, lane four, lane five, lane six, lane seven, lane eight, lane night, lane ten, lane eleven, and lane twelve are correctly classified. The classifier needed more clarity in predicting lane one until lane twelve category images. At higher speckle noise levels (), the classifier's performance is not satisfactory in predicting lane one until lane twelve image compared to lane two and three images.

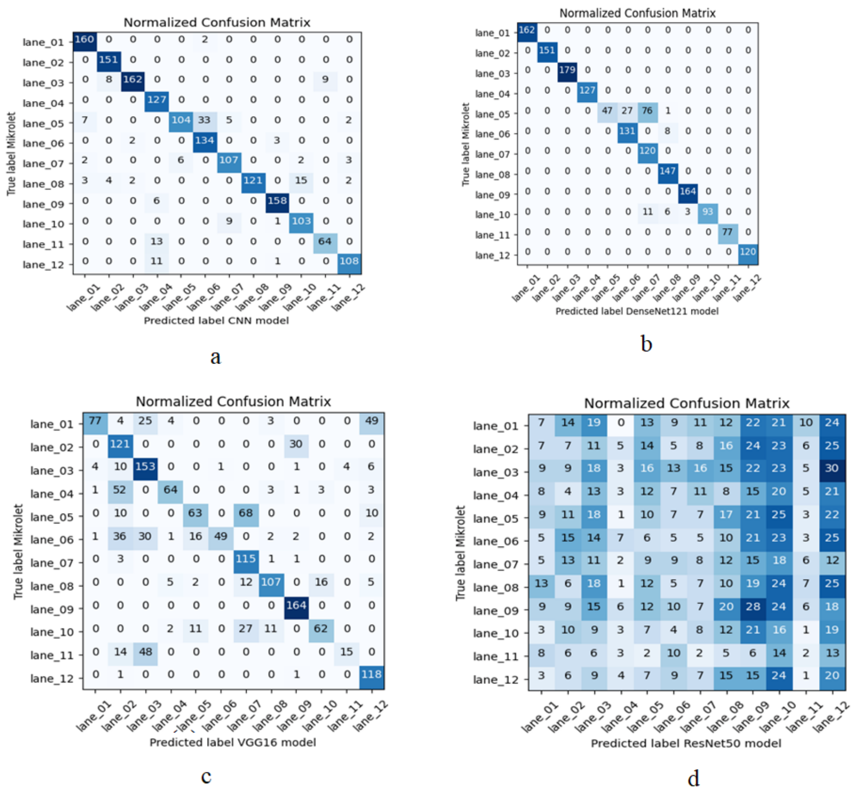

Fig.8 (a) Confusion Matrix CNN, (b) DenseNet121, (c) VGG16 and (d) ResNet50

The trained model was divided into two parts in our oriented CNN design. In the first phase, we have trained individual architectures with pre-trained unique models like VGG16, DenseNet121, ResNet50, and configurable architecture of the CNN model. In the second phase, we checked the results obtained from creating a single model and a CNN. The performance of our proposed method is compared with the existing methods. The performance analysis of four individual models is shown in Figure 9. Finally, we reached results with a classification of test accuracy of more than 100%. The accuracy of Individual performance for CNN, ResNet50, VGG16, and DenseNet121 is 100%, 66%, 100%, and 99%, respectively, and the image is accurately identified. The performance on this multiclass issue can be pictured utilizing the receiving characteristics (ROC) curve. The version of a classification model at all classification thresholds can be shown on the ROC curve graph. This curve outlines two parameters, such as the true-positive rate used to represent recall.

This model plans ROC curves for every disease, as demonstrated in Figure 10 (a,b,c,d). The ROC curve is plotted with the actual positive rate (TPR) against the false positive rate (FPR) on the x-axis. The performance results of the average value of the micro and macro middle ROC curve area showed that the performance of the CNN and Resnet models achieved 100% accuracy, and VGG16 achieved 96% accuracy. In comparison, Resnet only achieved 50% accuracy. (See Table 5). Comparison of the accuracy of the twelve classes' average micro and macro average RUC curve area values.

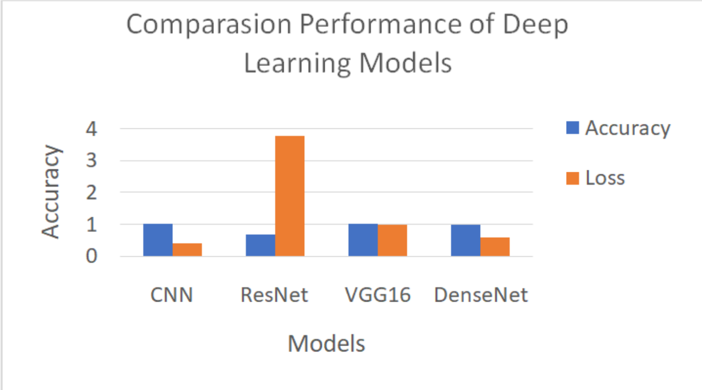

Figure 9. Graphical representation of comparative results

Table 4. The Precision, Recall, and F1-Score for four Deep Learning model

CNN				ResNet50			VGG16			DenseNet121		
Class	Prison	recall	f1-score	Prison	recall	f1-score	Prison	recall	f1-score	Prison	recall	f1-score
Lane_01	0.93	0.99	0.96	0.08	0.04	0.06	0.93	0.48	0.63	1.00	1.00	1.00
Lane_02	0.93	1.00	0.96	0.06	0.05	0.05	0.48	0.80	0.60	1.00	1.00	1.00
Lane_03	0.98	0.91	0.94	0.11	0.10	0.11	0.60	0.85	0.70	1.00	1.00	1.00
Lane_04	0.81	1.00	0.89	0.08	0.02	0.04	0.84	0.50	0.63	1.00	1.00	1.00
Lane_05	0.95	0.69	0.80	0.08	0.07	0.07	0.68	0.42	0.52	1.00	0.31	0.47
Lane_06	0.79	0.96	0.87	0.05	0.04	0.04	0.98	0.35	0.52	0.83	1.00	0.88
Lane_07	0.88	0.89	0.89	0.08	0.07	0.07	0.52	0.96	0.67	0.58	1.00	0.73
Lane_08	1.00	0.82	0.90	0.07	0.07	0.07	0.84	0.73	0.78	0.91	1.00	0.95
Lane_09	0.97	0.96	0.97	0.17	0.17	0.14	0.82	1.00	0.90	0.98	0.82	0.99
Lane_10	0.86	0.91	0.88	0.14	0.14	0.09	0.77	0.55	0.64	1.00	1.00	0.90
Lane_11	0.88	0.83	0.85	0.03	0.03	0.03	0.79	0.19	0.31	1.00	1.00	1.00
Lane_12	0.94	0.90	0.92	0.17	0.17	0.11	0.61	0.98	0.75	1.00	1.00	1.00

Table 5. Comparison Performance Micro and Macro average ROC curve area multiclass classification

Models	Micro Average ROC Curve area	Macro Average ROC curve area
CNN	99%	100%
VGG16	100%	100%
ResNet50	96%	96%
DenseNet121	49%	50%

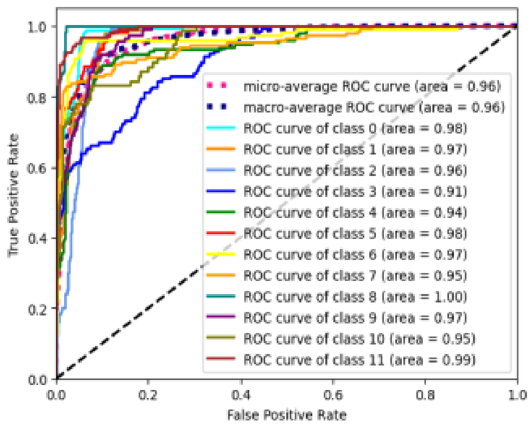

Fig.10 (a) Performance of micro and macro average ROC area CNN model

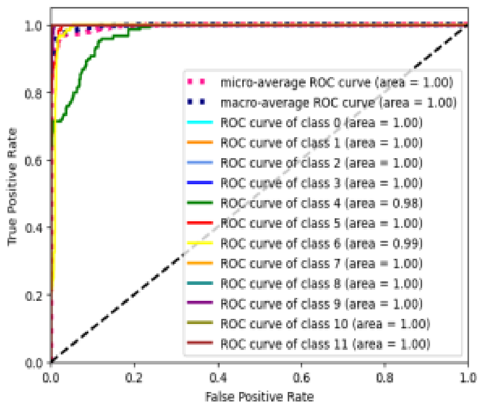

Fig.10 (b) Performance of micro and macro average ROC area DenseNet121 model

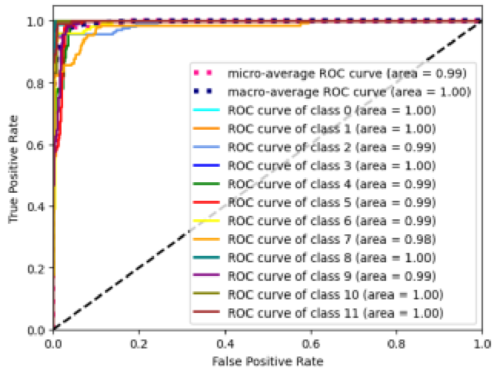

Fig.10 (c) Performance of micro and macro average ROC area VGG16 model

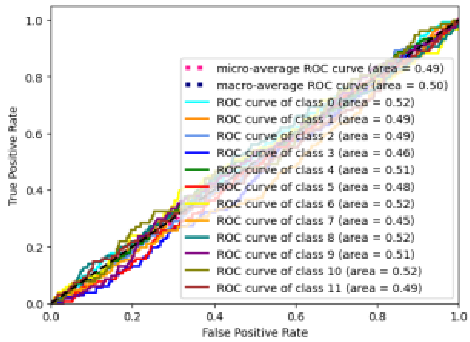

Fig.10 (d) Performance of micro and macro average ROC area ResNet50 model

4. CONCLUSIONS

This paper proposes the navigation system, an automatic multiclass classification of mikrolet vehicles using deep learning, which integrates different deep feature sets and several feature selections, and an optimized classifier with four deep learning models. The architecture is designed to classify small or medium datasets, a general and necessary task in real-world applications. However, it can be easily extended for large-scale data in the future. For a particular purpose, it is employed to identify and classify mikrolet vehicles in Dili city and compare deep learning models with the multiclass classification of the evaluation performance of each model. In the future, more complex conditions will be studied, such as traffic jams, lane-changing, and variations in traffic volume. Furthermore, we compare the computational complexity of these models and increase the dataset.

ACKNOWLEDGMENT

I want to thank JICA program of CADEFESH II to support this paper.

REFERENCES

[1] SUSANTY, Ade Pratiwi, et al. PelaksanaanTertibBerlalu Lintas TerhadapPemakaian Helm BagiMahasiswaPengendaraSepeda Motor di Dalam Area Kampus Universitas LancangKuningMenurutUndang-UndangNomor 22 Tahun 2009. Journal of Criminology and Justice, 2021, 1.1: 6-10.

[2] SYAFWANDI, Muhammad Isradi; DERMAWAN, Widodo Budi; ADITHYA, Randhy. Analysis Of Public Transport Service Performance (Case Study: Mikrotrans Jak 36 TrayekCilangkap-Cililitan).

[3] GUPTE, Surendra, et al. Detection and classification of vehicles. IEEE Transactions on intelligent transportation systems, 2002, 3.1: 37-47.

[4] LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. Nature, 2015, 521.7553: 436-444.

[5] CHEN, Wei; GUO, Fangzhou; WANG, Fei-Yue. A survey of traffic data visualization. IEEE Transactions on Intelligent Transportation Systems, 2015, 16.6: 2970-2984.

[6] BANINO, Andrea, et al. Vector-based navigation using grid-like representations in artificial agents. Nature, 2018, 557.7705: 429-433.

[7] HAM, Seung Woo, et al. Investigating the influential factors for practical application of multiclass vehicle detection for images from uncrewed aerial vehicles using deep learning models. Transportation Research Record, 2020, 2674.12: 553-567.

[8] THEISSLER, Andreas, et al. Confusion Vis: Comparative evaluation and selection of multiclass classifiers based on confusion matrices. Knowledge-Based Systems, 2022, 247: 108651.

[9] BJERRUM, EsbenJannik. SMILES enumeration as data augmentation for neural network modelling of molecules. arXiv preprint arXiv:1703.07076, 20

AUTHOR

I am **Abreu Andre Boavida**, From Timor-Leste, as an assistant lecturer at the National University of Timor Leste, Faculty of Engineering, Department of Informatics Engineering. I was a graduate undergraduate in Indonesia IST-Akprind Jogjakarta in 2013. I am studying master's at Gifu University-Japan, Department of Intelligent Science and Engineering, Informatics Division.