Vehicle Classification using Machine Learning Techniques

NADA FUAD KHATTAB[1], RAED ALAZAIDAH[2], ALA'A AL-SHAIKH[3],NIDAL YOUSSEF[1], NAWAF ALSHDAIFAT[4], MOHMMAD DMOUR[1]

[1]Faculty of IT, Department of Computer Science,
Zarqa University,
Zarqa,
JORDAN

[2]Faculty of IT, Department of and AI,
Zarqa University,
Zarqa,
JORDAN

[3]Faculty of IT
Department of Cyber Security
Zarqa University,
Zarqa,
JORDAN

[4]Faculty of IT,
Applied Science Private University,
Amman,
JORDAN

Abstract: - During the last few years, several real-life applications have attempted to utilize the proven high capabilities of artificial intelligence in general and machine learning in particular. Machine learning has been utilized in several domains, such as spam detection, image recognition, recommendation systems, self-driving cars, and medical diagnosis. This paper aims to survey the most related work of utilizing machine learning in vehicle classification. Moreover, the paper proposes a comparative analysis for identifying and determining the best classification model, best learning strategy, and the best feature selection method. Hence, four different vehicle datasets have been used to train seventeen classification models and five well-known feature selection methods with respect to several evaluation metrics such as Accuracy, True Positive ratio, Precision, and Recall. The results reveal that RandomForest and LMT are the best classifiers when it comes to handling vehicle datasets respectively. Considering the second objective, the Trees strategy showed the best performance.Furthermore,CorrelationAttributeEval,and ReliefFAttributeEval, are the best choices for handling the step of feature selection.

Key-Words: -Classification, Classifier, Feature Selection, Learning strategies, Machine learning, Vehicle classification.

Received: March 9, 2024. Revised: August 11, 2024. Accepted: September 8, 2024. Available online: October 4, 2024.

1 Introduction

Machine Learning (ML) has played an important role in the field of vehicle classification which refers to the process of identifying different types of vehicles and classifying them into one from a single view based on several input data such as images, videos,etc, [1]. Technology such as this is critical in industries that similarly cantered around concepts like transportation, security, and urban planning, [2]. Leveraging ML algorithms, such as convolutional neural networks (CNNs), those systems can go through vast amounts of data to properly classify the vehicles and significantly improve traffic management capabilities along with logistics and surveillance. In this research paper, we discussed the methods of ML and explained vehicle classification using ML where by discussing state-of-the-art and future directions in this rapidly growing area, [3], [4].

Artificial Intelligence (AI) encompasses any computer system capable of mimicking human cognitive functions to accomplish sophisticated objectives and pick up from experience. It could be anything from understanding natural language and recognizing patterns to the making of decisions, and solving problems independently, [5].

Moreover, Machine learning (ML): A subset of AI that uses algorithms to train models in order for it to give predictions, [6].

ML is an application of artificial intelligence (AI) that enables systems to automatically learn and improve from experience without being explicitly programmed. Thesemodels can accomplish with data, gaining accuracy on predictions by using it one time after another. The machine learning process involves the following key stages: Data collection,pre-processing, model training, evaluation, and deployment. ML is used in many fields such as medicine,healthcare, and more other areas with greater impacts on people's lives, [7]. Machine learning is used in these fields in image and speech recognition, natural language processing, recommendation systems, and predictive analytics, [8].

ML has four types: supervised, unsupervised,semi-supervised, and reinforcement, [9], [10]. Supervised learning, is a subcategory of machine learning and artificial intelligence, [11], [12]. It uses labelleddatasets to train models in order to be able to classify data in correctly. On the other hand,Unsupervised learning is a type of machine learning that learns from data without human follow up, in this type, the datasets are not labelledthe models can discover the patterns, [13]. Unlike supervised learning, unsupervised machine learning models are given the data, unlabelled, and allowed to discover patterns and insights without any human interaction, [14]. Semi-supervised learning is a machine learning approach that entails training a model using a dataset that includes both labelledand unlabeled data. Reinforcement learning is the third type of machine learning, [9], [15]. Here, models are self-trained on reward and punishment mechanisms, [16]. It's about taking the best possible action or path to gain maximum rewards and minimum punishment through observations in a specific situation. It acts as a signal to positive and negative behaviours, [17], [18], [19].

This research paper aims to achieve two main objectives: the first objective aims to identify the best classifier in vehicle classification. To achieve this objective, seventeen different classifiers that belong to the main strategies of machine learning have been used for evaluation. Eight evaluation metrics such as Accuracy, Precision, Recall, F-score, and others are used to evaluate the selected classifiers. There is a secondary goal linked to the first main goal, which is to choose the appropriate strategy from among the six used.

The second main objective of this paper is to identify the best feature selection method to be used in vehicle classification. To achieve that, five popular feature selection methods have been evaluated and compared using the same classifiers in the first objective with respect to three evaluation metrics, namely, Accuracy, Precision, and Recall.

The rest of the paper is organized as follows: Section 2 surveys some of the related works in the domain of utilizing ML techniques in vehicle classification. Section 3 presents the methodology, results, and discussion. Section 4 concludes and suggests future directions.

2 Related Work

In [20], autonomous driving and intelligent transportation have acknowledged the importance of vehicle positioning and classification technologies. The system used the SSD (Single Shot Multibox Detector) algorithm to accomplish vehicle classification and location. In the context of vehicle classification, a number of crucial steps—image collecting, picture calibration, model training, and model detection—were thoroughly explained. Using pre-labeled data was one of the strategies used in the annotation of photographs to increase annotation process efficiency. In recent years, the SSD algorithm—which is recognized for having a high degree of efficiency and Accuracy in target detection—has been widely used for tasks involving target location and classification recognition.

From my perspective, the SSD algorithm's effectiveness and precision in vehicle classification make it an appealing option for autonomous driving. However, its dependence on pre-labeled data poses the possibility of bias, and its ability to operate in real-time on embedded systems with limited resources has not been evaluated. Moreover, the impact of sensor noise, occlusions, and the system's ability to cope with different real-world environments was not analyzed. Hence, these limitations highlight the necessity of conducting more research to further improve this methodology and ensure its robustness when it is implemented in real autonomous driving scenarios.

In [21] road traffic accidents (RTAs) are a major issue with high fatality and injuries worldwide. In this study, the RTA modeling and analysis, sortingout and determiningtheir offset using machine learning classifiers were studied. In total, it checked seven other ways apart from naive Bayes to tackle the issue of missing data: logistic regression, k-nearest neighbor, AdaBoosting, support vector machines, and random forests. The study, which used a real-world RTA dataset from Gauteng province of South Africa claimed to provide guidance to policymakers and traffic authorities. Evaluation measures reference on receiver operating characteristic curves with tweaks, dimensionality reduction techniques, and performance indexessuch as Accuracy, Precision-Recall, and root mean square error. * Pragmatically, the best combination, as determined empirically in this paper (10), is the random forest classifier and multiple imputations via chained equations. This finding has direct implications for RTA modeling efforts.

I would like to suggest that, this study opens a new direction for discovering the efficacy of machine learning classifiers of road traffic accidents using Random Forest. Nevertheless, due to the regional nature of the study and the targets of events used, results deliver a useful foundation for further investigation and providerelevant recommendations on how to construct sophisticated predictive models or enhance strategies for increasing safety in transportation.

In [22], the vehicle counts and classification dataimportance to be comprehensively provided by the researchers in this study using ITS (Intelligent Transportation Systems). The researchers demonstrated a unique magnetometer-based real-time vehicle detection and classification system that operated out of the box without any additional computing hardware. In order to get large samples for training and validation, embedded pavement units collected data in a real-world setting. Examining magnetometer capabilities, nine vehicle classes were taken into consideration, surpassing similar approaches. In order to ensure low computing and memory requirements for real-time operation, classification used three-layer feedforward artificial neural networks (ANN) and creative time-domain waveform analysis for feature extraction. Research on sensor axe combinations with the goal of improving efficiency and minimizing classifier size. The system's strong classification efficiency on unknown samples was demonstrated by the results, which showed 74.67% with detection length and 73.73% without.

From my perspective, the research on the real-time online vehicle categorization system using a solitary tri-axial magnetic sensor is notable for its inventive and pragmatic methodology. The system's efficiency and flexibility are demonstrated by the utilization of an adaptive threshold-based algorithm, thorough real-world data collection, and successful implementation on a microcontroller. The study's future objectives, such as improving recognition efficiency and differentiating between different types of vehicles, demonstrate a proactive dedication to continuous enhancements. In my perspective, this research offers a hopeful option for intelligent transportation systems by combining creativity with a practical approach and establishing a clear path for future improvements.

In [23], the paper mentioned thatthe purpose of regulating traffic, smart traffic, and information systems are needed to gather traffic data from the appropriate sensors. In the last few years, security cameras have been placed to monitor and regulate traffic in this area. Numerous research projects have used image-processing techniques for traffic control in video surveillance systems. One example of an application for advanced cautioning or data extraction for real-time vehicle analysis was the video processing of traffic data captured by surveillance cameras. The literature on vehicle detection and classification methods was thoroughly reviewed in this work, which also included the unresolved issues in this field of study. It also examined a range of vehicle datasets that were employed in different research to assess the suggested methodologies.

In my point of view, this study effectively advocates for the significance of traffic data in intelligent transportation systems and underscores the possibilities of utilizing image processing techniques with security cameras. Nevertheless, the usefulness of the methodology might be enhanced by addressing unresolved obstacles such as occlusions and real-time constraints, conducting a thorough assessment of the datasets employed, and investigating other applications beyond basic vehicle detection to create a more complete and future-proof traffic monitoring method.

In [24], the authors talked about how the applications of vehicle detection in remote sensing photos in traffic, security, military, and surveillance have increased interest in the field. Convolutional Neural Networks (CNNs) were used in earlier studies, incorporating sophisticated methods such as homography augmentation, deep residual networks, multi-scale feature fusion, and hard example mining. Notably, researchers tackled low-resolution (LR) image identification issues as well as super-resolution (SR) issues in an integrated way. A Generative Adversarial Network (GAN) was used for unsupervised SR in order to get over the difficulty of gathering paired low-/high-resolution data. One unique tactic was to improve overall detection performance by applying back-propagating detection loss to the SR generator. The model outperformed cutting-edge techniques in deep learning and remote sensing, as shown by experimental results, making significant contributions to the discipline.

Regarding this sturdy, the technology achieves exceptional performance in vehicle detection for remote sensing, surpassing state-of-the-art techniques and reaching unprecedented levels.

In [25], convolutional neural networks (CNNs) have proven to be remarkably effective in fine-grained vehicle categorization in recent studies, particularly when it comes to detecting specific vehicle classifications. With our suggested channel max pooling (CMP) strategy, a new layer between fully connected and convolutional layers wasestablished, which is discriminative in nature of extracted features, unlike the usual back-propagation technique which prioritizes maximizing the loss function. By choosing maximum values, this CMP approach compressed feature maps into sub-groups. Notably, CNNs' efficiency was improved by the CMP layer's reduction of the amount of parameters. Experiments conducted on two fine-grained car datasets showed that CNNs enhanced with CMP greatly reduced parameters and increased classification Accuracy. In addition, CMP performed competitively when measured against cutting-edge techniques.

According to the study, the proposed CMP approach could be a promising way to complement and improve CNNs for fine-grained vehicle classification. These constraints could be addressed by conducting more extensive testing enhancing interpretability and mitigating overfitting; hence making it a valuable system to deploy in different real-world applications.

3 Research Methodology

This section explains in detail the steps followed in the study methodology, including the steps for data collection and description, it also includes the methodology in detail.

3.1 Description of Datasets

For this study, four distinct datasets were downloadedfrom the reliable source GitHub based on vehicle classification. these datasets all have information on the cars from different perspectiveswhich is important for classification,for example, the manufacturing city. Table 1 provides a summary of the main characteristics of each dataset.

Table 1. DatasetsDescription

Name	Instances	Attributes	Classes
Autos	26	206	6
Car1	406	8	3
Car2	261	8	3
Vehicle	846	19	4

3.2 Methodology

In pursuit of a powerful methodology for vehicle classification based on manufacturing city and other features, this study carefully executes a series of methodological steps. The initial step involves the careful selection of four inclusive datasets from GitHub sources. Following this, a thorough pre-processing stage ensues, handling missing values, outlier treatment, and normalization to ensure data integrity.

The heart of the methodology lies in the use of a diverse set of 17 classifiers, that belong to different learning strategies as described in the next section [19]. The classifiers are systematically trained with hyper parameter tuning and cross-validation, aiming to identify the most effective one for vehicle classification based on manufacturing city.

The comprehensive evaluation involves comparing performance metrics such as Accuracy, Precision, Recall, and TP ratescores. Through this detailed methodology, the research strives to uncover the optimal classifier for the nuanced task at hand. Figure 1 displaysthe main steps of the research methodology.

Fig. 1: Research Methodology

4 Evaluation Results

This section provides the results for the main two objectives of this research. Section 4.1 provides the results for the first objective, while Section 4.2 provides the results for the second objective.

4.1 Identifying the Best Classifier and the Best Learning Strategy

In the step of identifying the most effective classifier for vehicle classification based on manufacturing city and other features, this research employs several setsof 17 classifiers, each belongingto a specific learning strategy. The classifiers that are employed in this research are BayesNet, NaiveBayes, and NaiveBayesUpdateable from the Bayes strategy,and Logistic and MultilayerPerceptron from the functions strategy. Furthermore,Lazy learning strategies are embraced by IBK and KStar. Likewise, ASC, RandomCommittee, and RFC belong to the Meta strategy. DecisionTable, JRip, andPART follow rule-based learning. Finally, J48, LMT, Random Forest,and RandomTree from trees strategy. These classifiers, derived from the Weka (Waikato Environment for Knowledge Analysis) framework, are rigorously trained, fine-tuned, and evaluated using several metrics such as Accuracy, Precision, Recall, and true positive rate (TP), [25]. Accuracy is a metric that quantifies the extent to which a model is correct in its predictions. It is determined by dividing the number of accurately predicted cases by the total number of examples in the dataset. Precision is a measure of the accuracy of the positive predictions made by a model. It calculates the ratio of correctly predicted positive recordsaccording to the total predicted positives. Recall, often referred to as sensitivity or true positive rate, quantifies the capacity of a model to accurately detect all pertinent occurrences of the positive class. The calculation involves determining the proportion of correctly identified positive instances in relation to the combined number of correctly identified positive instances and incorrectly identified negative instances. The F1 Score is a quantitative measure that integrates Precision and Recall, yielding a harmonized assessment of both. It is especially advantageous in situations when there is an unequal distribution among the categories, and reaching a trade-off between accuracy and completeness is crucial. The following formulas detail how each metric is computed, providing a transparent and comprehensive approach to assessing the performance of the classifiers. Through this thorough evaluation process, the study aims to identify not only the best classifier but also the most effective learning strategy for the accurate task of vehicle classification based on manufacturing city and other features.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

$$\text{Precesion} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$\text{F1}_{\text{score}} = \frac{2*\text{Precesion}*\text{Recall}}{\text{Precesion+Recall}} \tag{4}$$

Table 2 summarizes that, among the vehicle classification classifiers, Random Forest whichbelongs to the Trees strategy, demonstrates strong overall performance with an Accuracy of 76.660% and notable Precision, Recall, and true positive rate values at 0.839, 0.840, and 0.839, respectively. The next best classifier is MultilayerPerceptron under the Functions strategy, excelling with the highest Accuracy at 80.000% and balanced Precision and Recall at 0.800. However, the worst classifiers, particularly in terms of Precision and Recall, are NaiveBayes and NaiveBayesUpdateable under the Bayes strategy, both exhibiting lower Accuracy at around 56.098%. In Table 2'PREC' stand for Precision 'RFC' stands for 'RandomizableFilteredClassifier', ASC stands for 'AttributeSelectedClassifier'.

Table 2. Comparative Analysis AMmongest the 17 Classifiers on Autos Dataset

Classifier	ACC	TP	PREC	Recall
BayesNet	68.293	0.683	0.694	0.683
NaiveBayes	**56.098**	**0.561**	**0.581**	**0.561**
NaiveBayesUpdateable	56.098	0.561	**0.581**	0.561
Average	**60.163**	**0.602**	**0.619**	**0.602**
Logistic	71.220	0.712	0.714	0.712
MultilayerPerceptron	80.000	0.800	0.804	0.800
Average	**75.610**	**0.756**	**0.759**	**0.756**
IBK	76.098	0.761	0.772	0.761
KStar	73.171	0.732	0.751	0.732
Average	**74.634**	**0.747**	**0.762**	**0.747**
ASC	78.049	0.780	0.786	0.780
RandomCommittee	**82.439**	0.824	0.825	0.824
RFC	62.439	0.624	0.639	0.624
Average	**74.472**	**0.745**	**0.756**	**0.745**
DecisionTable	65.366	0.654	0.742	0.654
JRip	73.171	0.732	0.731	0.732
PART	77.561	0.776	0.783	0.776
Average	**72.033**	**0.721**	**0.752**	**0.721**
J48	81.951	0.820	0.833	0.820
LMT	77.561	0.776	0.777	0.776
RandomForest	76.**660**	**0.839**	**0.840**	**0.839**
RandomTree	76.585	0.766	0.769	0.766
Average	**78.189**	**0.800**	**0.805**	**0.800**

In Table 3, the evaluation of vehicle classification classifiers based on various metrics is provided.

Table 3. Comparative Analysis AMmongest the 17 Classifiers on Car1 Datasets

Classifier	ACC	TP	PREC	Recall
BayesNet	67.816	0.135	0.728	0.678
NaiveBayes	70.881	0.112	0.766	0.709
NaiveBayesUpdateable	70.881	0.112	0.766	0.709
Average	**69.860**	**0.699**	**0.753**	**0.699**
Logistic	75.096	0.150	0.759	0.751
MultilayerPerceptron	75.862	0.143	0.759	0.759
Average	**75.479**	**0.755**	**0.759**	**0.755**
IBK	70.881	0.251	0.695	0.709
KStar	80.843	0.152	0.805	0.808
Average	**75.862**	**0.759**	**0.750**	**0.759**
ASC	78.544	0.152	0.777	0.785
RandomCommittee	84.674	0.132	0.843	0.847
RFC	75.479	0.187	0.754	0.755
Average	**79.566**	**0.796**	**0.791**	**0.796**
DecisionTable	71.648	0.287	0.686	0.716
JRip	75.479	0.159	0.758	0.755
PART	78.927	0.130	0.792	0.789
Average	**75.351**	**0.753**	**0.745**	**0.753**
J48	78.161	0.148	0.781	0.782
LMT	76.245	0.179	0.758	0.762
RandomForest	**86.207**	**0.102**	**0.862**	**0.862**
RandomTree	78.161	0.153	0.774	0.782
Average	**79.693**	**0.797**	**0.794**	**0.797**

Random Forest, under the Trees strategy, emerges as the best classifier. With a highAccuracy of 88.424%, it demonstrates noteworthy Precision, Recall, and true positive rate values at 0.884, 0.883, and 0.884, respectively. At its side, the next best-performing classifier is J48 under the Trees strategy, achieving an Accuracy of 87.931% and exhibiting high Precision, Recall, and true positive rate at 0.880. On the contrary, the lowest-performing classifiers, particularly in terms of Precision and Recall, include NaiveBayes and NaiveBayesUpdateable under the Bayes strategy, with an average Accuracy of 67.652%. These findings provide valuable insights into the comparative strengths and weaknesses of each classifier, aiding in informed decisions for vehicle classification tasks based on specific performance metrics.In Table 3 'RFC' stands for 'RandomizableFilteredClassifier', ASC stands for 'AttributeSelectedClassifier'.

Table 4 shows that Random Forest, applying the Trees strategy, is the best classifier based on the four metrics that were chosen. With a remarkable Accuracy of 86.207%, it exhibits amazing Recall, Precision, and true positive rate values of 0.862. KStar, which employs the Lazy method, is the next best classifier, coming in close second with an Accuracy of 80.843% with notable Precision, Recall, and true positive rates of 0.805 and 0.808. In contrast, the Bayes strategy's classifiers—NaiveBayes and NaiveBayesUpdateable—performed the worst, exhibiting decreased Precision and Recall values along with an average Accuracy of 69.860%.

Table 4. Comparative Analysis AMmongest the 17 Classifiers on Car2 Datasets

Classifier	ACC	TP	PREC	Recall
BayesNet	67.816	0.135	0.728	0.678
NaiveBayes	70.881	0.112	0.766	0.709
NaiveBayesUpdateable	70.881	0.112	0.766	0.709
Average	**69.860**	**0.699**	**0.753**	**0.699**
Logistic	75.096	0.150	0.759	0.751
MultilayerPerceptron	75.862	0.143	0.759	0.759
Average	**75.479**	**0.755**	**0.759**	**0.755**
IBK	70.881	0.251	0.695	0.709
KStar	80.843	0.152	0.805	0.808
Average	**75.862**	**0.759**	**0.750**	**0.759**
ASC	78.544	0.152	0.777	0.785
RandomCommittee	84.674	0.132	0.843	0.847
RFC	75.479	0.187	0.754	0.755
Average	**79.566**	**0.796**	**0.791**	**0.796**
DecisionTable	71.648	0.287	0.686	0.716
JRip	75.479	0.159	0.758	0.755
PART	78.927	0.130	0.792	0.789
Average	**75.351**	**0.753**	**0.745**	**0.753**
J48	78.161	0.148	0.781	0.782
LMT	76.245	0.179	0.758	0.762
RandomForest	**86.207**	**0.102**	**0.862**	**0.862**
RandomTree	78.161	0.153	0.774	0.782
Average	**79.693**	**0.797**	**0.794**	**0.797**

Based on the evaluation of vehicle classification classifiers in Table 5, LMT (Logistic Model Trees) using the Trees approach is the best performer with an Accuracy of 82.979%. Its noteworthy values of 0.827, 0.830, and 0.830 for Recall, Precision, and true positive rate, respectively, are demonstrated. Using the Functions method, MultilayerPerceptron closely follows, achieving a high Accuracy of 81.679% with notable values for Precision, Recall, and the true positive rate at 0.814 and 0.817. Conversely, classifiers using the Bayes strategy—NaiveBayes and NaiveBayesUpdateable, in particular—show worse overall performance, with a 49.882% average Accuracy as well as poorer Precision and Recall values. These updated insights give a clearer picture of the relative advantages and disadvantages of each classifier, enabling decision-making for jobs involving the classification of vehicles based on specific performance metrics. In Table 5, acc stands for 'Accuracy', 'Prec' for Precision, 'RFC' stands forRandomizableFilteredClassifier, 'ASC' stands for AttributeSelectedClassifier.

Table 5. Comparative Analysis AMmongest the 17 Classifiers on Vehicle Datasets

Classifier	ACC	TP	PREC	Recall
BayesNet	60.047	0.600	0.592	0.600
NaiveBayes	44.799	0.448	0.510	0.448
NaiveBayesUpdateable	44.799	0.448	0.510	0.448
Average	**49.882**	**0.499**	**0.537**	**0.499**
Logistic	79.787	0.798	0.797	0.798
MultilayerPerceptron	81.679	0.817	0.814	0.817
Average	**80.733**	**0.808**	**0.806**	**0.808**
IBK	69.858	0.699	0.691	0.699
KStar	71.395	0.714	0.701	0.714
Average	**70.627**	**0.707**	**0.696**	**0.707**
ASC	67.021	0.670	0.663	0.670
RandomCommittee	75.414	0.754	0.750	0.754
RFC	62.175	0.622	0.618	0.622
Average	**68.203**	**0.682**	**0.677**	**0.682**
DecisionTable	65.721	0.657	0.636	0.657
JRip	69.031	0.690	0.677	0.690
PART	71.513	0.715	0.713	0.715
Average	**68.755**	**0.687**	**0.675**	**0.687**
J48	72.459	0.725	0.722	0.725
LMT	**82.979**	**0.830**	**0.827**	**0.830**
RandomForest	76.005	0.760	0.752	0.760
RandomTree	70.922	0.709	0.712	0.709
Average	**75.591**	**0.756**	**0.753**	**0.756**

Table 6 summarizes the results in order to identify the best classifier that is suitable to use in predicting all selected datasets. By referring to Table 4 it is obvious that the best classifier is Random Forest. In Table 6 'RF' stands for 'Random Forest', 'RC' stands for 'RandomCommittee', and 'PREC' stands for 'Precision'

Table 6. Best Classifier with respect to Evaluation Metrics

Name	ACC	TP	PREC	Recall
Autos	RC	RF	RF	RF
Car1	RF	RF	RF	RF
Casr2	RF	RF	RF	RF
Vehicle	LMT	LMT	LMT	LMT

Table 7 summarizes the results in order to identify the best learning strategy that is used in this research. By referring to Table7, it is obvious that the best learning strategy is trees.

Table 7. Best Learning Strategy

Name	ACC	TP	PREC	Recall
Autos	**Meta**	Trees	Trees	Trees
Car1	Trees	Trees	Trees	Trees
Casr2	Trees	Trees	Trees	Trees
Vehicle	Trees	Trees	Trees	Trees

4.2 Identifying the Best Feature Selection Method

By choosing about half of the features from the four datasets, the research's second goal is to determine the best feature selection technique for improving classification performance. 'CAE' (ClassifierAttributeEval), 'CRE' (CorrelationAttributeEval), 'GR' (GainRatioAttributeEval), 'IG' (InfoGainAttributeEval), and 'RA' (ReliefFAttributeEval) are the five feature selection techniques that have been used to achieve this purpose. These techniques are essential for reducing the size of the dataset since they keep the most useful attributes while removing unnecessary or insignificant ones, [27], [28], [29], [30]. This goal is important because it aims to provide a feature subset that is well-balanced and efficient, which lowers dimensionality and improves the overall performance and Accuracy of the classification models, [31], [32]. Examining these various feature selection strategies will provide important information about how effective they are in comparison, which will help choose the best approach based on the particular features of the datasets being evaluated.

According to Table 8,the highest Accuracy values throughout our investigation were attained in large part because of the feature selection techniques. 'RA' (ReliefFAttributeEval) performed best with an Accuracy of about 87.805% in the Lazy strategy using the KStar classifier on the Autos dataset, whereas 'IG' (InfoGainAttributeEval) performed best with an Accuracy of 88.293%. These remarkable Accuracy values highlight how well these feature selection techniques work to dramatically improve the KStar classifier's classification Accuracy when used in combination with the Lazy strategy. This demonstrate the importance of several methods of feature selection, specifically 'RA' and 'IG', to maximize the KStar classifier's classification performance within the parameters of our particular classification job on the Autos dataset when considering Accuracy.

In this study, the effect of feature selection on improving the precisionmetric appears clearly visible in the results, as shown in Table 9. the highest precision values throughout thisstudy were attained due to the use of feature selection techniques.'RA' (ReliefFAttributeEval) performed best with a Precision of about 0.879% in the Lazy strategy using the KStar classifier on the Autos dataset, whereas 'IG' (InfoGainAttributeEval) performed best with a precision of 0.885. These remarkable Accuracy values highlight how well these feature selection techniques work to dramatically improve the KStar classifier's classification Accuracy when used in combination with the Lazy strategy. These results show how important feature selection is, specifically 'RA' and 'IG', to maximize the KStar classifier's classification performance within the parameters of our particular classification job on the Autos dataset when considering Precision.

Table 8. Evaluation of the Considered Feature Selection Methods on Auto Dataset with Respect to Accuracy Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	68.293	63.902	68.781	71.707	76.585
NaiveBayes	54.146	53.171	54.146	59.512	62.927
NaiveBayesUpdateable	54.146	53.171	54.146	59.512	62.927
Average	**58.862**	**56.748**	**59.024**	**63.577**	**67.480**
Logistic	72.683	56.098	73.659	74.146	72.195
MultilayerPerceptron	82.927	69.268	82.927	80.000	82.927
Average	**77.805**	**62.683**	**78.293**	**77.073**	77.561
IBK	82.439	75.122	82.439	81.463	80.000
KStar	71.707	66.829	71.220	**88.293**	**87.805**
Average	**77.073**	**70.976**	**76.829**	**84.878**	83.902
ASC	78.049	76.098	78.049	78.049	78.049
RandomCommittee	**87.805**	82.439	84.390	85.366	84.390
RFC	76.585	68.293	66.829	75.122	76.098
Average	**80.813**	**75.610**	**76.423**	**79.512**	79.512
DecisionTable	66.342	65.366	66.342	67.805	66.829
JRip	70.732	73.171	75.122	75.122	78.537
PART	74.146	73.171	77.561	72.195	77.073
Average	**70.407**	**70.569**	**73.008**	**71.707**	74.146
J48	78.049	76.098	82.439	76.098	77.561
LMT	82.927	75.610	80.976	80.000	79.024
RandomForest	84.390	**83.415**	**85.854**	85.366	86.342
RandomTree	78.049	75.122	79.512	83.902	83.902
Average	**80.854**	**77.561**	**82.195**	**81.341**	**81.707**

Table 9. Evaluation of the Considered Feature Selection Methods on Auto Dataset with Respect to Precision Metric

Classifier	***CAE***	***CRE***	***GR***	***IG***	***RA***
BayesNet	0.687	0.643	0.691	0.714	0.768
NaiveBayes	0.578	0.563	0.600	0.614	0.618
NaiveBayesUpdateable	0.578	0.563	0.600	0.614	0.618
Average	**0.614**	**0.590**	**0.630**	**0.647**	**0.668**
Logistic	0.732	0.552	0.745	0.767	0.737
MultilayerPerceptron	0.834	0.687	0.832	0.804	0.834
Average	**0.783**	**0.620**	**0.789**	**0.786**	**0.786**
IBK	0.828	0.754	0.825	0.818	0.805
KStar	0.734	0.678	0.736	**0.885**	**0.879**
Average	**0.781**	**0.716**	**0.781**	**0.852**	**0.842**
ASC	0.782	0.769	0.782	0.786	0.786
RandomCommittee	**0.878**	0.827	0.847	0.855	0.845
RFC	0.777	0.692	0.673	0.759	0.764
Average	**0.812**	**0.763**	**0.767**	**0.800**	**0.798**
DecisionTable	0.754	0.718	0.777	0.773	0.766
JRip	0.709	0.741	0.748	0.756	0.789
PART	0.744	0.735	0.787	0.716	0.772
Average	**0.736**	**0.731**	**0.771**	**0.748**	**0.776**
J48	0.781	0.765	0.834	0.765	0.777
LMT	0.832	0.759	0.812	0.803	0.793
RandomForest	0.846	**0.841**	**0.859**	0.855	0.864
RandomTree	0.785	0.756	0.808	0.841	0.840
Average	**0.811**	**0.780**	**0.828**	**0.816**	**0.819**

Table 10 shows the evaluation results for the considered feature selection methods with respect to the Recall metric on the Auto dataset.

According to Table 10, the best Recallhas been achieved when using IG,RA by KStar classifier that belongs to lazy strategy on Auto dataset.Also, comparing the results for Recall between the case when using all features as in Table 2, and the case of using 50 % of the features as in Table 10, it is clear that the performance has improved.

According to Table 11, the best Accuracy has been achieved when using CRE,GR, IG, and RA by RandomForest classifier that belongs to the trees strategy on the Car1 dataset.

Also, comparing the results for Accuracy between the case when using all features as in Table 3, and the case of using 50 % of the features as in Table 11, it is clear that the performance has improved.

Table 10. Evaluation of the Considered Feature Selection Methods on Auto Dataset with Respect to Recall Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.683	0.639	0.688	0.717	0.766
NaiveBayes	0.541	0.532	0.541	0.595	0.629
NaiveBayesUpdateable	0.541	0.532	0.541	0.595	0.629
Average	**0.588**	**0.568**	**0.590**	**0.636**	**0.675**
Logistic	0.727	0.561	0.737	0.741	0.722
MultilayerPerceptron	0.829	0.693	0.829	0.800	0.829
Average	**0.778**	**0.627**	**0.783**	**0.771**	**0.776**
IBK	0.824	0.751	0.824	0.815	0.800
KStar	0.717	0.668	0.712	**0.883**	**0.878**
Average	**0.771**	**0.710**	**0.768**	**0.849**	**0.839**
ASC	0.780	0.761	0.780	0.780	0.780
RandomCommittee	**0.878**	0.824	0.844	0.854	0.844
RFC	0.766	0.683	0.668	0.751	0.761
Average	**0.808**	**0.756**	**0.764**	**0.795**	**0.795**
DecisionTable	0.663	0.654	0.663	0.678	0.668
JRip	0.707	0.732	0.751	0.751	0.785
PART	0.741	0.732	0.776	0.722	0.771
Average	**0.704**	**0.706**	**0.730**	**0.717**	**0.741**
J48	0.780	0.761	0.824	0.761	0.776
LMT	0.829	0.756	0.810	0.800	0.790
RandomForest	0.844	**0.834**	**0.859**	0.854	0.863
RandomTree	0.780	0.751	0.795	0.839	0.839
Average	**0.808**	**0.776**	**0.822**	**0.814**	**0.817**

According to Table 12, the best Precisionhas been achieved when using CRE,GR, IG, and RA by RandomForest classifier that belongs to the trees strategy on the Precision Car1 dataset.

Also, comparing the results for Precisionbetween the case when using all features as in Table 3, and the case of using 50 % of the features as in Table 12, it is clear that the performance has improved.

Table 11. Evaluation of the Considered Feature Selection Methods on Car1 Dataset with Respect to Accuracy Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	66.749	66.503	66.503	66.503	66.503
NaiveBayes	66.256	65.764	65.764	65.764	65.764
NaiveBayesUpdateable	66.256	65.764	65.764	65.764	65.764
Average	**66.420**	**66.010**	**66.010**	**66.010**	**66.010**
Logistic	74.138	74.877	75.616	75.616	75.616
MultilayerPerceptron	75.616	73.399	75.616	75.616	75.616
Average	**74.877**	**74.138**	**75.616**	**75.616**	**75.616**
IBK	75.616	76.601	82.512	82.512	82.512
KStar	80.296	82.020	82.512	82.512	82.512
Average	**77.956**	**79.310**	**165.025**	**165.025**	**165.025**
ASC	83.744	83.251	83.744	83.744	83.744
RandomCommittee	**86.700**	85.468	86.453	86.453	86.453
RFC	75.862	79.064	81.527	81.527	81.527
Average	**82.102**	**82.594**	**83.908**	**83.908**	**83.908**
DecisionTable	74.631	74.384	74.877	74.877	74.877
JRip	77.586	77.094	80.296	80.296	80.296
PART	82.020	86.453	84.237	84.237	84.237
Average	**78.079**	**79.310**	**79.803**	**79.803**	**79.803**
J48	84.483	85.468	84.729	84.729	84.729
LMT	82.759	82.759	83.990	83.990	83.990
RandomForest	86.453	**88.424**	**87.685**	**87.685**	**87.685**
RandomTree	85.714	82.759	85.222	85.222	85.222
Average	**84.852**	**84.852**	**85.406**	**85.406**	**85.406**

Table 12. Evaluation of the Considered Feature Selection Methods on Car1 Dataset with Respect to Precision Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.732	0.710	0.719	0.719	0.719
NaiveBayes	0.744	0.719	0.733	0.733	0.733
NaiveBayesUpdateable	0.744	0.719	0.733	0.733	0.733
Average	**0.740**	**0.716**	**2.185**	**2.185**	**2.185**
Logistic	0.743	0.755	0.759	0.759	0.759
MultilayerPerceptron	0.767	0.759	0.793	0.793	0.793
Average	**0.755**	**0.757**	**0.776**	**0.776**	**0.776**
IBK	0.756	0.768	0.823	0.823	0.823
KStar	0.804	0.824	0.835	0.835	0.835
Average	**0.780**	**0.796**	**1.658**	**1.658**	**1.658**
ASC	0.839	0.827	0.839	0.839	0.839
RandomCommittee	**0.862**	0.850	0.860	0.860	0.860
RFC	0.755	0.792	0.815	0.815	0.815
Average	**0.819**	**0.823**	**0.838**	**0.838**	**0.838**
DecisionTable	0.716	0.718	0.721	0.721	0.721
JRip	0.763	0.763	0.798	0.798	0.798
PART	0.815	0.865	0.842	0.842	0.842
Average	**0.765**	**0.782**	**0.787**	**0.787**	**0.787**
J48	0.841	0.851	0.845	0.845	0.845
LMT	0.828	0.820	0.838	0.838	0.838
RandomForest	0.860	**0.881**	**0.873**	**0.873**	**0.873**
RandomTree	0.853	0.822	0.851	0.851	0.851
Average	**0.846**	**0.844**	**0.852**	**0.852**	**0.852**

According to Table 13, the best Recall has been achieved when using CRE,GR, IG, and RA by RandomForest classifier that belongs to the trees strategy on the RecallCar1 dataset.

According to Table 14, the best Accuracy has been achieved when using CRE,GR, IG, and RA by Random Forest classifier that belongs to the trees strategy on the AccuracyCar2 dataset. Also, comparing the results for Accuracy between the case when using all features as in Table 4, and the case of using 50 % of the features as in Table 14, it is clear that the performance has improved.

Table 13. Evaluation of the Considered Feature Selection Methods on Car1 Dataset with Respect to Recall Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.667	0.665	0.665	0.665	0.665
NaiveBayes	0.663	0.658	0.658	0.658	0.658
NaiveBayesUpdateable	0.663	0.658	0.658	0.658	0.658
Average	**0.664**	**0.660**	**1.981**	**1.981**	**1.981**
Logistic	0.741	0.749	0.756	0.756	0.756
MultilayerPerceptron	0.756	0.734	0.756	0.756	0.756
Average	**0.749**	**0.742**	**0.756**	**0.756**	**0.756**
IBK	0.756	0.766	0.825	0.825	0.825
KStar	0.803	0.820	0.825	0.825	0.825
Average	**0.780**	**0.793**	**1.650**	**1.650**	**1.650**
ASC	0.837	0.833	0.837	0.837	0.837
RandomCommittee	**0.867**	0.855	0.865	0.865	0.865
RFC	0.759	0.791	0.815	0.815	0.815
Average	**0.821**	**0.826**	**0.839**	**0.839**	**0.839**
DecisionTable	0.746	0.744	0.749	0.749	0.749
JRip	0.776	0.771	0.803	0.803	0.803
PART	0.820	0.865	0.842	0.842	0.842
Average	**0.781**	**0.793**	**0.798**	**0.798**	**0.798**
J48	0.845	0.855	0.847	0.847	0.847
LMT	0.828	0.828	0.840	0.840	0.840
RandomForest	0.865	**0.884**	**0.877**	**0.877**	**0.877**
RandomTree	0.857	0.828	0.852	0.852	0.852
Average	**0.849**	**0.849**	**0.854**	**0.854**	**0.854**

Moreover, according to Table 15, the best Precision has been achieved when using CRE, GR, IG, and RA by RandomForest classifier that belongs to the tree's strategy on the AccuracyCar2 dataset. Also, comparing the results for Precision between the case when using all features as in Table 4, and the case of using 50 % of the features as in Table 15, it is clear that the performance has been improved.

Table 14. Evaluation of the Considered Feature Selection Methods on Car2 Dataset with Respect to Accuracy Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	65.134	68.199	68.199	69.349	63.985
NaiveBayes	69.732	66.284	66.284	67.433	66.667
NaiveBayesUpdateable	69.732	66.284	66.284	67.433	66.667
Average	**68.199**	**66.922**	**66.922**	**68.072**	**65.773**
Logistic	72.414	76.628	76.628	72.797	76.245
MultilayerPerceptron	71.648	72.797	72.797	73.946	72.031
Average	**72.031**	**74.713**	**74.713**	**73.372**	**74.138**
IBK	77.395	75.862	75.862	77.395	81.226
KStar	78.544	80.460	80.460	81.226	84.291
Average	**77.969**	**78.161**	**78.161**	**79.310**	**82.759**
ASC	76.245	77.395	77.395	74.713	77.778
RandomCommittee	84.291	81.609	81.609	83.142	84.674
RFC	79.310	76.245	76.245	75.862	81.992
Average	**79.949**	**78.416**	**78.416**	**77.906**	**81.481**
DecisionTable	68.966	71.264	71.264	70.881	70.115
JRip	75.479	76.245	76.245	78.927	74.330
PART	75.479	75.096	75.096	77.778	74.713
Average	**73.308**	**74.202**	**74.202**	**75.862**	**73.052**
J48	77.395	78.161	78.161	84.291	80.077
LMT	76.245	78.161	78.161	78.927	83.525
RandomForest	**84.674**	**82.759**	**82.759**	**85.058**	**85.441**
RandomTree	83.908	79.694	79.694	81.226	85.058
Average	**80.556**	**79.693**	**79.693**	**82.376**	**83.525**

Table 15. Evaluation of the Considered Feature Selection Methods on the Car2 Dataset with Respect to Precision Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.718	0.730	0.730	0.746	0.700
NaiveBayes	0.752	0.723	0.723	0.734	0.729
NaiveBayesUpdateable	0.752	0.723	0.723	0.734	0.729
Average	**0.741**	**0.725**	**0.725**	**0.738**	**0.719**
Logistic	0.736	0.778	0.778	0.742	0.774
MultilayerPerceptron	0.732	0.735	0.735	0.753	0.736
Average	**0.734**	**0.757**	**0.757**	**0.748**	**0.755**
IBK	0.770	0.766	0.766	0.773	0.809
KStar	0.789	0.815	0.815	0.818	0.851
Average	**0.780**	**0.791**	**0.791**	**0.796**	**0.830**
ASC	0.759	0.772	0.772	0.747	0.777
RandomCommittee	0.842	0.809	0.809	0.830	0.843
RFC	0.842	0.769	0.769	0.760	0.817
Average	**0.814**	**0.783**	**0.783**	**0.779**	**0.812**
DecisionTable	0.640	0.687	0.687	0.684	0.655
JRip	0.756	0.778	0.778	0.794	0.727
PART	0.742	0.761	0.761	0.778	0.757
Average	**0.713**	**0.742**	**0.742**	**0.752**	**0.713**
J48	0.778	0.780	0.780	0.849	0.816
LMT	0.773	0.788	0.788	0.788	0.842
RandomForest	**0.849**	**0.827**	**0.827**	**0.849**	**0.854**
RandomTree	0.834	0.796	0.796	0.807	0.846
Average	**0.809**	**0.798**	**0.798**	**0.823**	**0.840**

Table 16. Evaluation of the Considered Feature Selection Methods on the Car2 Dataset with Respect to Recall Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.651	0.682	0.682	0.693	0.640
NaiveBayes	0.697	0.663	0.663	0.674	0.667
NaiveBayesUpdateable	0.697	0.663	0.663	0.674	0.667
Average	**0.682**	**0.669**	**0.669**	**0.680**	**0.658**
Logistic	0.724	0.766	0.766	0.728	0.762
MultilayerPerceptron	0.716	0.728	0.728	0.739	0.720
Average	**0.720**	**0.747**	**0.747**	**0.734**	**0.741**
IBK	0.774	0.759	0.759	0.774	0.812
KStar	0.785	0.805	0.805	0.812	0.843
Average	**0.780**	**0.782**	**0.782**	**0.793**	**0.828**
ASC	0.762	0.774	0.774	0.747	0.778
RandomCommittee	0.843	0.816	0.816	0.831	0.847
RFC	0.843	0.762	0.762	0.759	0.820
Average	**0.816**	**0.784**	**0.784**	**0.779**	**0.815**
DecisionTable	0.690	0.713	0.713	0.709	0.701
JRip	0.755	0.762	0.762	0.789	0.743
PART	0.755	0.751	0.751	0.778	0.747
Average	**0.733**	**0.742**	**0.742**	**0.759**	**0.730**
J48	0.774	0.782	0.782	0.843	0.801
LMT	0.762	0.782	0.782	0.789	0.835
RandomForest	**0.847**	**0.828**	**0.828**	**0.851**	**0.854**
RandomTree	0.839	0.797	0.797	0.812	0.851
Average	**0.806**	**0.797**	**0.797**	**0.824**	**0.835**

Moreover, according to Table 16, the best Recallhas been achieved when using CRE, GR, IG, and RA by RandomForest classifier that belongs to the trees strategy on the AccuracyCar2 dataset.

Furthermore, according to Table 17, the best Accuracy has been achieved when using CAE, RA by LMT classifier that belongs to the tree's strategy on the Accuracy on Vehicle dataset.

Furthermore, according to Table 18, the best Precision has been achieved when using CAE, and RA by the LMT classifier that belongs to the tree's strategy on the Precision on Vehicle dataset.

Furthermore, according to Table 19, the best Recall has been achieved when using CAE, RA by LMT classifier that belongs to the tree's strategy on the Recall on Vehicle dataset.

Table 17. Evaluation of the Considered Feature Selection Methods on Vehicle Dataset with Respect to Accuracy Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	60.166	56.856	60.875	60.875	58.511
NaiveBayes	45.745	42.080	43.617	43.617	39.835
NaiveBayesUpdateable	45.745	42.080	43.617	43.617	39.835
Average	**50.552**	**47.006**	**49.370**	**49.370**	**46.060**
Logistic	70.804	67.849	70.567	70.567	68.676
MultilayerPerceptron	74.114	72.459	72.104	72.104	72.931
Average	**72.459**	**70.154**	**71.336**	**71.336**	**70.804**
IBK	70.686	69.385	68.440	68.440	67.612
KStar	70.804	69.385	72.577	72.577	69.385
Average	**70.745**	**69.385**	**70.508**	**70.508**	**68.499**
ASC	65.957	66.430	66.194	66.194	58.629
RandomCommittee	71.040	72.222	69.858	69.858	69.504
RFC	67.731	63.357	64.894	64.894	63.830
Average	**68.243**	**67.337**	**66.982**	**66.982**	**63.987**
DecisionTable	65.957	63.475	64.184	64.184	63.475
JRip	64.894	63.712	69.858	69.858	62.530
PART	70.567	68.558	68.440	68.440	69.385
Average	**67.139**	**65.248**	**67.494**	**67.494**	**65.130**
J48	70.213	**73.050**	68.322	68.322	70.686
LMT	**76.596**	72.931	**73.759**	**73.759**	**74.705**
RandomForest	73.286	72.340	72.813	72.813	71.158
RandomTree	70.922	65.721	71.040	71.040	66.785
Average	**72.754**	**71.011**	**71.483**	**71.483**	**70.833**

Table 18. Evaluation of the Considered Feature Selection Methods on Vehicle Dataset with Respect to Precision Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.602	0.569	0.609	0.609	0.585
NaiveBayes	0.457	0.421	0.436	0.436	0.398
NaiveBayesUpdateable	0.457	0.421	0.436	0.436	0.398
Average	**0.505**	**0.470**	**0.494**	**0.494**	**0.460**
Logistic	0.708	0.678	0.706	0.706	0.687
MultilayerPerceptron	0.741	0.725	0.721	0.721	0.729
Average	**0.725**	**0.702**	**0.714**	**0.714**	**0.708**
IBK	0.707	0.694	0.684	0.684	0.676
KStar	0.708	0.694	0.726	0.726	0.694
Average	**0.708**	**0.694**	**0.705**	**0.705**	**0.685**
ASC	0.660	0.664	0.662	0.662	0.586
RandomCommittee	0.710	0.722	0.699	0.699	0.695
RFC	0.677	0.634	0.649	0.649	0.638
Average	**0.682**	**0.673**	**0.670**	**0.670**	**0.640**
DecisionTable	0.660	0.635	0.642	0.642	0.635
JRip	0.649	0.637	0.699	0.699	0.625
PART	0.706	0.686	0.684	0.684	0.694
Average	**0.672**	**0.653**	**0.675**	**0.675**	**0.651**
J48	0.702	**0.730**	0.683	0.683	0.707
LMT	**0.766**	0.729	**0.738**	**0.738**	**0.747**
RandomForest	0.733	0.723	0.728	0.728	0.712
RandomTree	0.709	0.657	0.710	0.710	0.668
Average	**0.728**	**0.710**	**0.715**	**0.715**	**0.709**

Table 19. Evaluation of the Considered Feature Selection Methods on Vehicle Dataset with Respect to Recall Metric

Classifier	CAE	CRE	GR	IG	RA
BayesNet	0.580	0.556	0.593	0.593	0.585
NaiveBayes	0.554	0.380	0.451	0.451	0.373
NaiveBayesUpdateable	0.554	0.380	0.451	0.451	0.373
Average	**0.563**	**0.439**	**0.498**	**0.498**	**0.444**
Logistic	0.698	0.667	0.695	0.695	0.675
MultilayerPerceptron	0.735	0.713	0.717	0.717	0.727
Average	**0.717**	**0.690**	**0.706**	**0.706**	**0.701**
IBK	0.700	0.689	0.687	0.687	0.672
KStar	0.695	0.681	0.711	0.711	0.682
Average	**0.698**	**0.685**	**0.699**	**0.699**	**0.677**
ASC	0.640	0.654	0.638	0.638	0.575
RandomCommittee	0.700	0.715	0.697	0.697	0.692
RFC	0.669	0.627	0.650	0.650	0.635
Average	**0.670**	**0.665**	**0.662**	**0.662**	**0.634**
DecisionTable	0.646	0.619	0.620	0.620	0.613
JRip	0.644	0.620	0.687	0.687	0.633
PART	0.704	0.684	0.680	0.680	0.689
Average	**0.665**	**0.641**	**0.662**	**0.662**	**0.645**
J48	0.694	**0.720**	0.677	0.677	0.701
LMT	**0.762**	0.719	**0.732**	**0.732**	**0.742**
RandomForest	0.718	0.711	0.713	0.713	0.699
RandomTree	0.701	0.664	0.710	0.710	0.675
Average	**0.719**	**0.704**	**0.708**	**0.708**	**0.704**

5 Conclusion and Future Work

In this paper, two main objectives have been achieved. The first is the identification of the best classifier that suits the domain of vehicle classification and the identification of the best learning strategy. The second is the identification of the best feature selection method in order to reduce the dimensionality of the datasets and thus to improve the performance. Regarding the first objective, two classifiers showed the best results: RandomForest and LMT. Considering the second objective, the Trees strategy showed the best performance. According to the third objective, CorrelationAttributeEval on car1, ReliefFAttributeEval on the Car2 dataset, ClassifierAttributeEval on the vehicle dataset, and InfoGainAttributeEval on the Autos dataset showed the best performance. Hence, it is highly recommended to consider an ensemble model that consists of the two best classifiers to solve the problem of vehicle classification as a future work. Metaheuristic algorithms can be used in the future to design feature selection algorithms with greater performance.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the first author used QuillBot AI Paraphrasing tool, in order to paraphrase few paragraphs. After using this tool, the first author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References:

[1] L. Zhuo, L. Jiang, Z. Zhu, J. Li, J. Zhang, and H. Long, "Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks," *Machine Vision and Applications*, vol. 28, pp. 793-802, (2017).

[2] M. Won, "Intelligent traffic monitoring systems for vehicle classification: A survey" *IEEE Access*, *8*, 73340-73358, (2020).

[3] Q. Xue, K. Wang, J. J. Lu, and Y. Liu, "Rapid driving style recognition in car-following using machine learning and vehicle trajectory data," *Journal of Advanced Transportation*, 2019.

[4] E. S., Hasan, M. K. Hassan, R. Saeed. "Machine learning technologies for secure vehicular communication in the internet of vehicles: recent advances and applications". *Security and Communication Networks*, 2021, 1-23, (2021).

[5] N. Nasir, A. Kansal, O. Alshaltone, F. Barneih, M. Sameer, A. Shanableh. "Water quality classification using machine learning algorithms." *Journal of Water Process Engineering*, *48*, 102920, (2022).

[6] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, & A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science"*, 3-21 (2020).

[7] I. H. Sarker. "Machine learning: Algorithms, real-world applications, and research directions." *SN computer science*, *2*(3), 160, 2021.

[8] L. Y. Yab, N. Wahid, and R. A. Hamid, "A Meta-Analysis Survey on the Usage of Meta-Heuristic Algorithms for Feature Selection on High-Dimensional Datasets," *IEEE Access*, vol. 10, pp. 122832-122856, 2022.

[9] H. Liang, X. Zhang, X. Hong, Z. Zhang, M. Li, G. Hu, and F. Hou, "Reinforcement learning enabled dynamic resource allocation in the Internet of Vehicles," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 7, pp. 4957-4967, (2020).

[10] Z. Salah, K. Salah & E. Elsoud, (2024). Spatial domain noise removal filtering for low-resolution digital images. *Indonesian Journal of Electrical Engineering and Computer Science*, *34*(3), 1627-1642.

[11] Z. Salah, K. Salah & E. Elsoud, (2024). Spatial domain noise removal filtering for low-resolution digital images. *Indonesian Journal of Electrical Engineering and Computer Science*, *34*(3), 1627-1642.

[12] T. Jiang, J. L. Gradus, & A. J. Rosellini. "Supervised machine learning: a brief primer". *Behavior Therapy*, *51*(5), 675-687 (2020).

[13] M. Alzyoud, R. Alazaidah, M. Aljaidi, G. Samara, M. Qasem, M. Khalid, &N. Al-Shanableh, (2024). Diagnosing diabetes mellitus using machine learning techniques. *International Journal of Data and Network Science*, 8(1), 179-188.

[14] R. Alazaidah, G. Samara, S. Almatarneh, M. Hassan, M. Aljaidi, H. & Mansur, (2023). Multi-Label Classification Based on Associations. *Applied Sciences*,*13*(8),5081.

[15] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 273-292, 2019.

[16] C. Zhang, Y. Lu. "Study on artificial intelligence: The state of the art and future prospects". *Journal of Industrial Information Integration*, 23, 100224, (2021).

[17] Z. Ma, D. Chang, J. Xie, Y. Ding, S. Wen, X. Li, ,J. Guo, "Fine-grained vehicle classification with channel max pooling modified CNNs," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3224-3233, 2019.

[18] R. C. Jisha, J. M. Amrita, A. R. Vijay, and G. S. Indhu, "Mobile app recommendation system using machine learning classification," *in 2020 Fourth international conference on computing methodologies and communication (ICCMC),* March 2020, pp. 940-943.

[19] R. Alazaidah and M. A. Almaiah, "Associative classification in multi-label classification: An investigative study," *Jordanian Journal of Computers and Information Technology*, vol. 7, no. 2, 2021.

[20] Y. Yang, "Realization of Vehicle Classification System Based on Deep Learning," in 2020 IEEE International

Conference on Power, *Intelligent Computing and Systems (ICPICS)*, July 2020, pp. 308-311.

[21] T. Bokaba, W. Doorsamy, and B. S. Paul, "Comparative study of machine learning classifiers for modelling road traffic accidents," *Applied Sciences*, vol. 12, no. 2, p. 828, (2022.)

[22] P. Sarcevic, S. Pletl, and A. Odry, "Real-time vehicle classification system using a single magnetometer," *Sensors*, vol. 22, no. 23, pp. 9299, 2022.

[23] V. K. Kiran, P. Parida, and S. Dash, "Vehicle detection and classification: a review," in International Conference on Innovations in *Bio-Inspired Computing and Applications*, August 2020, pp. 45-56.

[24] M. Cao, H. Ji, Z. Gao, and T. Mei, "Vehicle detection in remote sensing images using deep neural networks and multi-task learning," ISPRS Annals of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, vol. 2, pp. 797-804, 2020.

[25] M. I. Pavel, S. Y. Tan, and A. Abdullah, "Vision-based autonomous vehicle systems based on deep learning: A systematic literature review," *Applied Sciences*, vol. 12, no. 14, pp. 6831, 2022.

[26] A.Ghaben; M. Anbar; I. H.Hasbullah; and S. Karuppayah. Mathematical Approach as Qualitative Metrics of Distributed Denial of Service Attack Detection Mechanisms. In September 2021, date of current version September 13, 2021. *IEEE Access.* Digital Object Identifier 10.1109/ACCESS.2021.3110586.

[27] H. A. Owida, H. S. Migdadi, O. S. M. Hemied, N. F. F. Alshdaifat, S. F. A. Abuowaida, & R. S. Alkhawaldeh, (2022). Deep learning algorithms to improve COVID-19 classification based on CT images. *Bulletin of Electrical Engineering and Informatics*, 11(5), 2876-2885.

[28] H. A. Owida, B. A. H. Moh'd, N. Turab, J. Al-Nabulsi, &S. Abuowaida, (2023). The Evolution and Reliability of Machine Learning Techniques for Oncology. *International Journal of Online & Biomedical Engineering*, 19(8).

[29] Samara, Ghassan. "Intelligent reputation system for safety messages in VANET." *Int J ArtifIntell* 9, no. 3 (2020): 439-447.

[30] A. Y. Alhusenat, H. A. Owida, H. A. Rababah, J. Al-Nabulsi, &S. Abuowaida, (2023). A Secured Multi-Stages Authentication Protocol for IoT Devices. *Mathematical Modelling of Engineering Problems*, 10(4).

[31] G. Samara, (2020, November). Wireless sensor network MAC energy-efficiency protocols: a survey. *In 2020 21st International Arab Conference on Information Technology* (ACIT) (pp. 1-5). IEEE.

[32] E. Elbasi, &A. I. Zreikat, (2023). Heart Disease Classification for Early Diagnosis based on Adaptive Hoeffding Tree Algorithm in IoMTData. *International Arab Journal of Information Technology*, *20*(1), 38-48.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research has been funded by Zarqa University.

Conflict of Interest

The authors have no conflicts of interest to declare.