iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54971

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Material Selection in Electric Vehicle Chassis: A Machine Learning Approach

Santhosh Kumar R[1], Harshit V[2], Sathya Selvaraj Sinnasamy[3]
[1, 2]*UG Student,* [3]*Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, India*

Abstract: ***This research project focuses on applying machine learning (ML) algorithms for material selection in Electric Vehicle (EV) chassis design. By leveraging ML techniques, we aim to identify suitable materials based on key mechanical properties, enhancing structural performance, safety, and efficiency. Our evaluation of various ML models provides valuable insights for accurate material selection, contributing to the advancement of electric vehicle technology. By considering these evaluation metrics, we gain insights into the effectiveness of the ML models and their potential for accurate material selection in the context of EV chassis design.***

Keywords: Electrical Vehicle, Elastic Modulus, Machine learning, Support Vector Machines, Confusion Matrix

I. INTRODUCTION

The design and development of an Electric Vehicle (EV) chassis demand careful consideration of material selection to ensure optimal structural performance, safety, and efficiency. Traditionally, material selection processes have relied on empirical knowledge and expert judgment. However, with the advent of Machine Learning (ML) techniques, data-driven approaches have emerged as powerful tools for supporting informed decision-making in various fields, including material engineering. In this research project, we aim to leverage ML algorithms to assist in the selection of suitable materials for EV chassis design. Specifically, we focus on utilizing a limited set of seven features, including ID, Ultimate Tensile Strength, Yield Strength, Young's Modulus, Shear Modulus, Poisson's Ratio, and Density. These features capture important mechanical properties of machine design materials that directly influence the performance and durability of the EV chassis.

It is worth noting that due to missing data in the original dataset, only these seven features were considered for analysis. To address this limitation, appropriate data pre-processing techniques, such as imputation or exclusion, were applied to ensure the reliability and completeness of the dataset. This refined dataset, enriched with the selected features forms the basis for training and assessing machine learning models.

In the pursuit of identifying the most accurate and reliable ML model for material selection, various classification algorithms were employed. These models include logistic regression, support vector machines (SVM), k-nearest neighbours (KNN), naive Bayes, and ensemble models combining SVM and KNN. The application of diverse models allows for a comprehensive comparison and evaluation of their performance in predicting material suitability for EV chassis design. We used evaluation metrics to assess the performance of each ML model. These metrics provide quantitative measures of how well the models classify materials based on the selected features. By considering these evaluation metrics, we gain insights into the effectiveness of the ML models and their potential for accurate material selection in the context of EV chassis design.

Moreover, the design and development of an EV chassis involve complex considerations beyond material selection. Factors such as aerodynamics, energy efficiency, and crashworthiness are also critical for ensuring the overall performance and safety of the vehicle. While this research project focuses specifically on material selection, it should be noted that integrating ML techniques into other aspects of EV chassis design holds promising avenues for future exploration. Additionally, it is important to acknowledge the dynamic nature of the automotive industry and the continuous advancements in material engineering. As new materials with improved properties and characteristics emerge, it will be necessary to update and refine the ML models used for material selection. This ongoing research effort ensures that the models remain relevant and effective in an ever-evolving technological landscape.

Furthermore, the success of ML algorithms in material selection heavily depends on the quality and representativeness of the underlying dataset. Obtaining comprehensive and diverse data on material properties is essential for training accurate and robust models. Collaborations with material manufacturers, industry experts, and academic institutions can help gather extensive datasets and further enhance the applicability and reliability of ML-driven material selection processes.

In conclusion, by harnessing the power of ML algorithms, this research project aims to revolutionize the material selection process for EV chassis design. The utilization of data-driven approaches, the careful consideration of relevant features, and the evaluation of multiple ML models contribute to the advancement of high-performance and sustainable electric vehicles. Through continued research and collaboration, the integration of ML techniques in automotive design can shape the future of transportation, creating greener, safer, and more efficient vehicles.

II. METHODOLOGY

In this section, we shall delve into the methodologies employed in our research endeavors. It consists of collecting the data, pre-processing, extracting features, selecting the important features and training the model with the machine learning algorithms.

A. Data Collection

In the data collection phase of our research, we focused on assembling a comprehensive dataset specifically tailored to the mechanical properties of electric vehicle (EV) chassis materials. The dataset was obtained from the Autodesk Material Library, a reputable source known for its rich collection of materials and associated properties. The dataset is publicly accessible on Kaggle. The dataset comprises 15 distinct features, each representing a different mechanical property of the EV chassis materials.

The properties include parameters such as Poisson's ratio, Young's modulus, elongation at break, yield strength, and other key indicators. These properties are crucial for assessing the suitability and performance of materials in the construction of electric vehicle chassis. By including a broad range of properties, we aim to provide a comprehensive understanding of the characteristics and behavior of various materials under different conditions.

In total, the dataset contains 1552 rows and 15 columns in which, each row corresponding to a unique set of mechanical properties for an EV chassis material. This extensive collection of data enables researchers and engineers to explore, analyze, and draw meaningful insights. All the features and their description of the dataset is shown in the below Table 1.

TABLE I
DATASET EXPLANATION

S.No	Features	Description
1	Std	Standard (Std) of the material
2	ID	Unique Identification code for the Material (ID)
3	Material Name	Name of the Material
4	Heat Treatment	Method used in Heat Treatment
5	Su	Ultimate Tensile Strength (Su) in Mpa
6	Sy	Yield Strength (Sy) in Mpa
7	A5	Elongation at Break or Strain (A5) in Percentage
8	Bhn	Brinell Hardness Number (BHN) in Microhardness Units
9	E	Elastic Modulus (E) Mpa
10	G	Shear Modulus (G) in Mpa
11	Mu	Poisson's Ratio (mu)
12	Ro	Density (Ro) Kg/m3
13	Ph	Pressure at Yield (pH) in Mpa
14	VH	Vickers Hardness Number in HV

B. Data Pre-Processing

In the data pre-processing phase, an essential aspect of data preparation, various operations are performed on the raw dataset to make it suitable for subsequent data processing procedures. This section outlines the specific steps we undertook to pre-process the dataset and extract relevant features for our research. Initially, we focused on removing any rows that contained missing values or duplicated entries. This step ensures that the dataset remains consistent and reliable, as missing or duplicate data can introduce biases and inaccuracies in subsequent analyses. Next, we examined the correlation matrix of the dataset shown in the Figure 1 to identify any low correlated features.

To generate the data correlation matrix, one can easily utilize the 'corr()' function in the pandas library on the analyzed data frame [9]. From our observations, we decided to remove the following columns: 'Std', 'ID', 'Heat treatment', 'Desc', 'A5', 'Bhn', 'pH', and 'HV'. By eliminating these features, we aimed to reduce the complexity of the dataset while retaining the most informative attributes for our analysis.

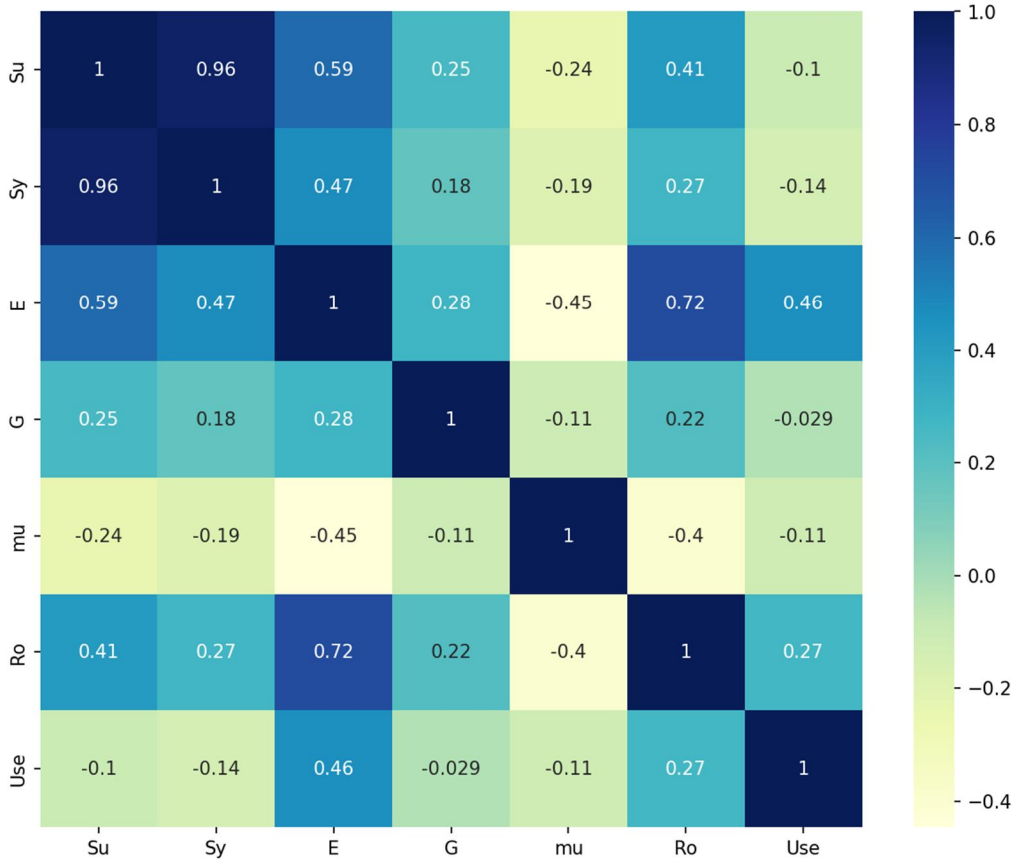


Fig.1 Correlation Matrix

Furthermore, during our examination of the dataset, we discovered an imbalance in the distribution of categories within the 'Use' variable. Imbalanced data distributions can lead to biased learning models, resulting in lower predictive accuracy. To mitigate this issue, we employed a technique known as data balancing. In particular, we addressed the imbalanced nature of the dataset by both undersampling and oversampling. The Figure 2 below shows how the dataset is distributed before balancing, with respect to the 'Use' variable.

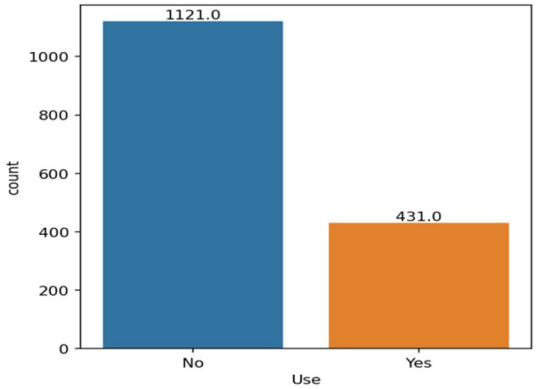


Fig.2 Dataset distribution before balancing

Random undersampling is a form of data sampling process of reducing the number of samples that are in majority class instances [4]. Using a python library imbalanced learn, we will be using the combination of over- and under-sampling. First we undersampled the dataset by 85%, ensuring that the majority classes were adequately represented without overwhelming the minority classes and oversampling by 15%. The Figure 3 below shows how the dataset is distributed after balancing, with respect to the 'Use' variable.

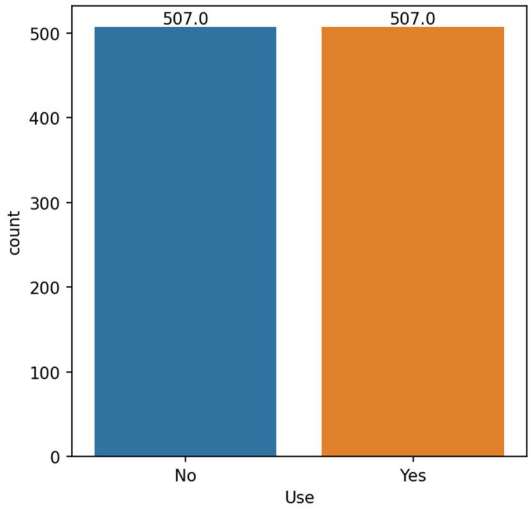


Fig.3 Dataset distribution after balancing

Conversely, oversampling involves by randomly duplicate the instances in the minority classes to balance the dataset further [4]. This technique enables the learning model to gain more exposure to the underrepresented classes, enhancing its ability to make accurate predictions. By employing a combination of undersampling and oversampling methods, we aimed to achieve the best performance and mitigate the negative effects of class imbalance.

C. Feature Extraction

Feature extraction plays a vital role in data processing by converting raw data into new characteristics that preserve the fundamental information found in the initial dataset. In our research, we performed feature extraction using specific features such as 'Su', 'Sy', 'E', 'G', 'mu', and 'Ro'. From these features, we derived a new feature called 'Use', which serves as a categorical variable indicating whether a material for an EV chassis is usable or not (Yes/No). This newly extracted feature will act as the independent variable in our dataset. In Figure 4, we provide an overview of how the 'Use' feature is extracted from the mentioned features using Python's pandas library.

```
df['Use'] = (
    (df['Su'].between(292, 683)) &
    (df['Sy'].between(212, 494)) &
    (df['E'].between(196650, 217350)) &
    (df['G'].between(47400, 110600)) &
    (df['mu'].between(0.225, 0.375)) &
    (df['Ro'].between(6288, 9432))
).map({True: 'Yes', False: 'No'})
```

Fig.4 Extracting 'Use' Feature using Pandas Library

To carry out the feature extraction process, we employed the widely used pandas library in Python. This library offers extensive functionality for data manipulation and analysis, making it an excellent choice for extracting features from our dataset. If all the conditions are met, we assign the value 'Yes' to the 'Use' feature. Conversely, if any of the conditions are not satisfied, we assign the value 'No'. Through this feature extraction process, we enhance the utility of our dataset, enabling us to explore the relationship between material properties and the usability of EV chassis materials. These insights play a crucial role in guiding material selection decisions and optimizing the design of electric vehicle chassis, ultimately contributing to the development of more efficient and reliable electric vehicles.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

D. Feature Selection

In the feature selection phase of our research, we aim to reduce irrelevant features for our model by carefully choosing only the relevant data and eliminating noise within the dataset. The selection of appropriate features is essential for building a robust and efficient model that focuses on the most significant factors influencing the outcome.

For our research, we have selected the useful features from the dataset using correlation matrix. The features, 'Su', 'Sy', 'E', 'G', 'Mu' and 'Ro' has been selected as dependent features shown in Table 2 and 'Use' feature that was extracted will be our independent feature shown in Table 3 to serve as the variables in our analysis. The dependent features represent the properties that we want to predict or explain, while the independent features act as the predictors or inputs to our model.

TABLE II
DEPENDENT FEATURES

S.No	Features	Description
1	Su (Ultimate Tensile Strength)	Numerical
2	Sy (Yield Strength)	Numerical
3	E (Elastic Modulus)	Numerical
4	G (Shear Modulus)	Numerical
5	Mu (Poisson's Ratio)	Numerical
6	Ro (Density)	Numerical

TABLE III
INDEPENDENT FEATURES

S.No	Features	Description
1	Use (Usable: Yes/No)	Categorical

E. Training the Model

In model training, we will discuss into the process of training the machine learning models using our dataset. To achieve this, we have selected several machine learning algorithms like Logistic Regression (LR), Support Vector Machines (SVM), Naive Bayes, and Ensemble Learning as these are the popular algorithms used among the researchers, when it comes to material classification [1], [2], [3], [12]. The goal is to utilize these algorithms to train the data and classify whether a material is usable or not.

To initiate the training process, we follow a general procedure that involves the following steps:

1) The dependent variable is the outcome or response variable that we aim to predict or explain in a study. It is influenced by or dependent on the independent variable(s), which are the factors or variables that we manipulate or observe to understand their impact on the dependent variable.

2) The dataset is divided into two parts, with 80% allocated to the training set and 20% assigned to the testing set, ensuring a balanced distribution.

3) The training set is trained by the model, where it learns patterns and connections between the independent variables and the dependent variable.

4) After training, the model predicts the outcomes using the testing set. The performance of the model is then evaluated by comparing its predictions to the actual values in the testing set, allowing us to assess how well the model generalizes to unseen data.

a) *Logistic Regression:* It is a mathematical method helps in analysing the data and predict outcomes with limited possibilities, such as binary yes or no outcomes. While it is typically applied to numerical data, logistic regression can also handle discrete variables by incorporating both continuous and discrete predictors. [10].

b) *Support Vector Machines:* In Machine Learning, an SVM model represents samples as points in a feature space. It aims to separate samples of different categories using a hyperplane positioned to maximize the distance from the nearest samples of each category [1], [12].

c) *Naïve Bayes Classifier:* It is a practical Bayesian learning method used for data with multiple attributes. It assumes the independence of attributes given the class and calculates the conditional probability of a class based on its priori probability and likelihood from the training data. [10].

d) *K-Nearest Neighbors:* KNN is a classification technique that predicts outcomes for test data by comparing them to the K nearest neighbors. It uses Euclidean distance to measure the proximity between data points and is considered a nonparametric model [11].

e) *Ensemble Learning:* Ensemble Learning involves combining predictions from multiple base estimators to enhance generalization and robustness compared to a single estimator. In our model, we will leverage a combination of SVM and KNN algorithms to improve overall performance [5].

III.RESULTS AND DISCUSSIONS

In the initial stage of our analysis, we generated individual confusion matrices for each classification model. A confusion matrix can be represented as a table that is used in classification models. It is used to evaluate predicted and actual class labels. The most common labels are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). When the values are positive and the model also identifies it as positive, then it is called true positive. When the values are negative and the model identifies it as positive then it is called false negative [5]. We use the confusion matrix evaluate the performance of the model. The attributes such as accuracy, precision and recall can be evaluated from the confusion matrix.

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

$$\text{Precision} = \frac{TP}{TP+FP}$$

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

$$\text{F- Measure} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

Fig.5 Attributes of confusion matrix

In our study, we evaluated the model performance of five classification models: logistic regression, support vector machine (SVM), k-nearest neighbours (KNN), naive Bayes, and an ensemble model combining SVM and KNN [6].

Logistic regression exhibited an accuracy of 91%, indicating that it correctly classified 91% of the instances in the dataset. It achieved a precision of 86% and recall of 100% for predicting the positive class. The F1-score was 0.92. The model demonstrated a high true positive rate (TPR) of 1.0, meaning it accurately identified all positive instances. The false positive rate (FPR) of 0.18 suggests a relatively low incidence of misclassifying instances.

TABLE IV
CLASSIFICATION REPORT FOR LOGISTIC REGRESSION

	Precision	**Recall**	**F1-Score**
No	0.98	0.88	0.93
Yes	0.90	0.99	0.94
Accurate			0.94
Micro average	0.94	0.93	0.94
Weighted average	0.94	0.94	0.94
No	0.98	0.88	0.93

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

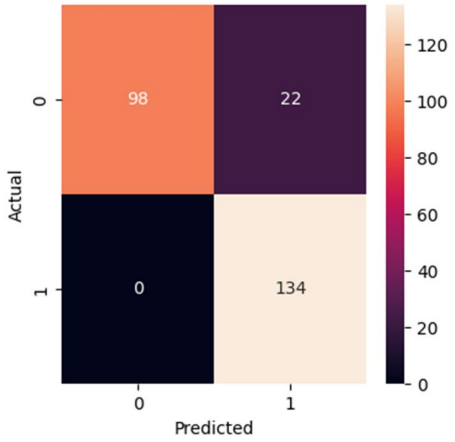


Fig.6 Logistic Regression Confusion Matrix

Support Vector Machines achieved a 90% accuracy on the dataset. It displayed a precision of 84% and recall of 100%, yielding an F1-score of 0.91. The TPR indicated accurate identification of positive instances at a rate of 1.0. However, the model showcased a higher FPR of 0.21 compared to logistic regression, suggesting a relatively higher rate of misclassifying negative instances as positive.

TABLE V
CLASSIFICATION REPORT FOR SVM

	Precision	Recall	F1-Score
No	1.00	0.86	0.92
Yes	0.89	1.00	0.94
Accurate			0.93
Micro average	0.94	0.93	0.93
Weighted average	0.94	0.93	0.93
No	1.00	0.86	0.92

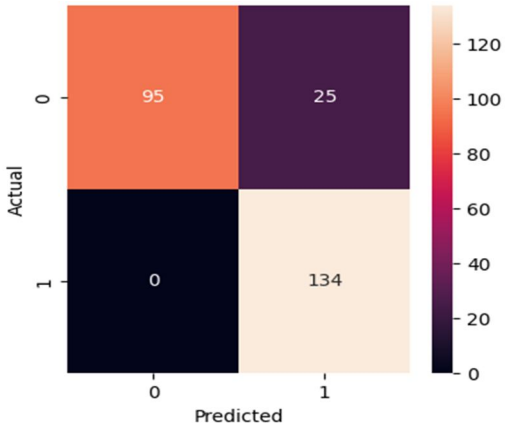


Fig.7 SVM Confusion Matrix

KNN surpassed the other models with an impressive of accuracy of 96%. It achieved a precision of 94% and recall of 99%, resulting in an impressive F1-score of 0.96. The model demonstrated a high TPR of 0.99, indicating accurate identification of positive instances. Upon comparing the classification models' performances, KNN emerged as the top-performing model in terms of accuracy, precision, recall, and F1-score.

It showcased exceptional classification capabilities, with high accuracy and a well-balanced precision-recall trade-off. Logistic regression also delivered commendable results, demonstrating a balanced precision and recall. SVM, though slightly lower in accuracy compared to logistic regression and KNN, performed well with high recall. Naive Bayes exhibited lower precision but achieved high recall.

TABLE VI
CLASSIFICATION REPORT FOR KNN

	Precision	**Recall**	**F1-Score**
No	0.99	0.9	0.97
Yes	0.96	0.99	0.98
Accurate			0.98
Micro average	0.98	0.98	0.98
Weighted average	0.98	0.98	0.98
No	0.99	0.9	0.97

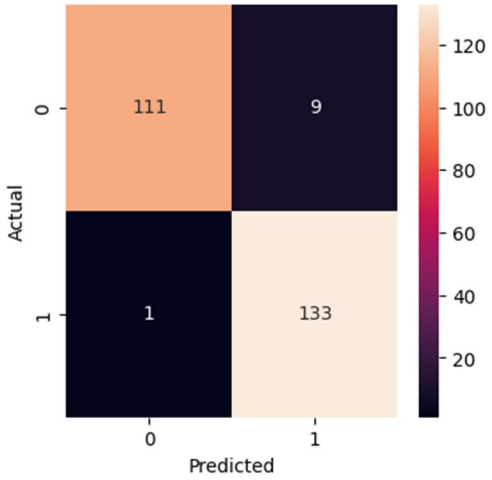


Fig.8 KNN Confusion Matrix

Naive Bayes achieved an accuracy of 86% on the dataset. It exhibited a precision of 80% and recall of 99%, resulting in an F1-score of 0.88. The TPR was 0.99, indicating accurate identification of positive instances. However, the model had a higher FPR of 0.28 compared to other models, suggesting a relatively higher rate of incorrectly classifying negative instances as positive.

TABLE VII
CLASSIFICATION REPORT FOR NAÏVE BAYES

	Precision	**Recall**	**F1-Score**
No	0.99	0.72	0.83
Yes	0.80	0.99	0.88
Accurate			0.86
Micro average	0.89	0.85	0.86
Weighted average	0.89	0.86	0.86
No	0.99	0.72	0.83

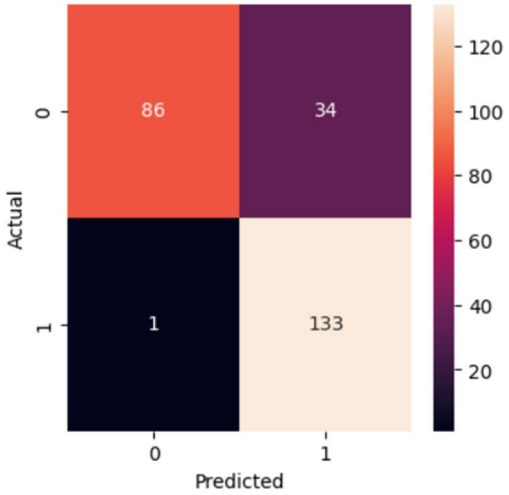


Fig.9 Naïve Bayes Confusion Matrix

The ensemble model, combining SVM and KNN, achieved an accuracy of 95%. It exhibited a precision of 91% and recall of 100%, resulting in an F1-score of 0.95. Similar to SVM, the ensemble model had a relatively higher FPR compared to logistic regression and KNN.

TABLE VIII
CLASSIFICATION REPORT FOR ENSEMBLE LEARNING

	Precision	**Recall**	**F1-Score**
No	1.00	0.53	0.69
Yes	0.70	1.00	0.82
Accurate			0.78
Micro average	0.85	0.76	0.76
Weighted average	0.84	0.78	0.76
No	1.00	0.53	0.69

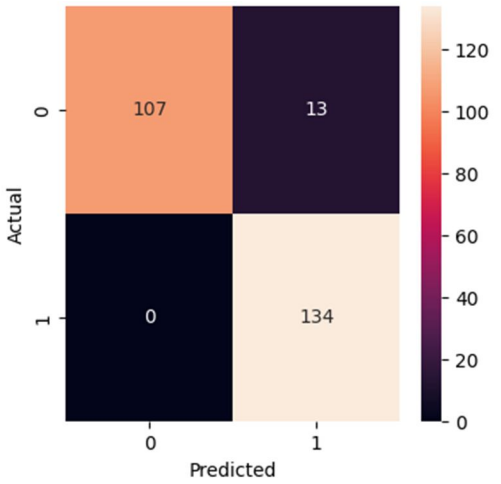


Fig.10 Ensemble Learning Confusion Matrix

While we evaluate the classification models we use various attributes such as accuracy, precision, recall and F1-score. Upon comparing the performance of various models, it was determined that the KNN (K-Nearest Neighbours) model outperformed others across these metrics, establishing it as the leading model. KNN demonstrated superior classification capabilities by achieving high accuracy, which measures the model's correctness. This indicates that KNN was able to correctly classify significant portion of the test data.

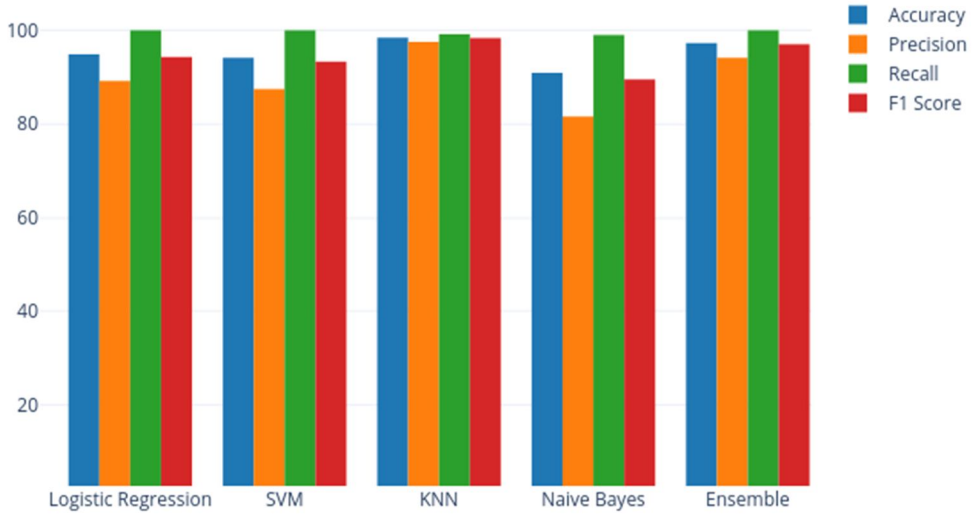


Fig.11 Bar Plot for various models

Additionally, KNN effectively balanced precision and recall Precision is an assessing metric used in the ratio of accurately predicted positive instances to the total instances predicted as positive. We can evaluate the model's accuracy and sensitivity by calculating the predicted instances and the actual instances. It emphasizes the model's capability to correctly identify positive instances. The balanced precision and recall scores indicate that KNN was able to both accurately identify positive instances and avoid false positives. Another model that showed commendable performance was logistic regression. Moreover, it attained balanced precision and recall scores, signifying its proficiency in accurately classifying positive instances while minimizing the occurrence of false positives. Although not explicitly stated, it can be inferred that logistic regression had a high accuracy as well, although slightly lower than that of KNN.

TABLE IX
COMPARING MODEL'S ACCURACY

S.No	Model	Accuracy
1	Logistic Regression	0.948
2	SVM	0.940
3	KNN	0.984
4	Naïve Bayes	0.909
5	Ensemble	0.972

IV. CONCLUSION

KNN emerged as the best-performing model with high accuracy, precision, recall, and a well-balanced precision-recall trade-off. Logistic regression also performed well with balanced precision and recall. SVM and the ensemble model showed high recall but had higher false positive rates [7]. Naive Bayes achieved high recall but lower precision. Overall, KNN demonstrated exceptional classification capabilities, but the choice of model should consider specific dataset characteristics and research objectives [8].

REFERENCES

[1] Rashidi, A., Sigari, M. H., Maghiar, M., & Citrin, D. (2016). An analogy between various machine-learning techniques for detecting construction materials in digital images. KSCE Journal of Civil Engineering, 20, 1178-1188. [2] J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[2] Alaloul, W. S., & Qureshi, A. H. (2021). Material classification via machine learning techniques: construction projects progress monitoring. In Deep Learning Applications. IntechOpen.

[3] Penumuru, D. P., Muthuswamy, S., & Karumbu, P. (2020). Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. Journal of Intelligent Manufacturing, 31(5), 1229-1241.

[4] Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research, 18(1), 559-563.

[5] Marom, N. D., Rokach, L., & Shmilovici, A. (2010, November). Using the confusion matrix for improving ensemble classifiers. In 2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel (pp. 000555-000559). IEEE.

[6] Salmon, B. P., Kleynhans, W., Schwegmann, C. P., & Olivier, J. C. (2015, July). Proper comparison among methods using a confusion matrix. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3057-3060). IEEE.

[7] Meyer, K. B., & Pauker, S. G. (1987). Screening for HIV: can we afford the false positive rate?. New England journal of medicine, 317(4), 238-241.

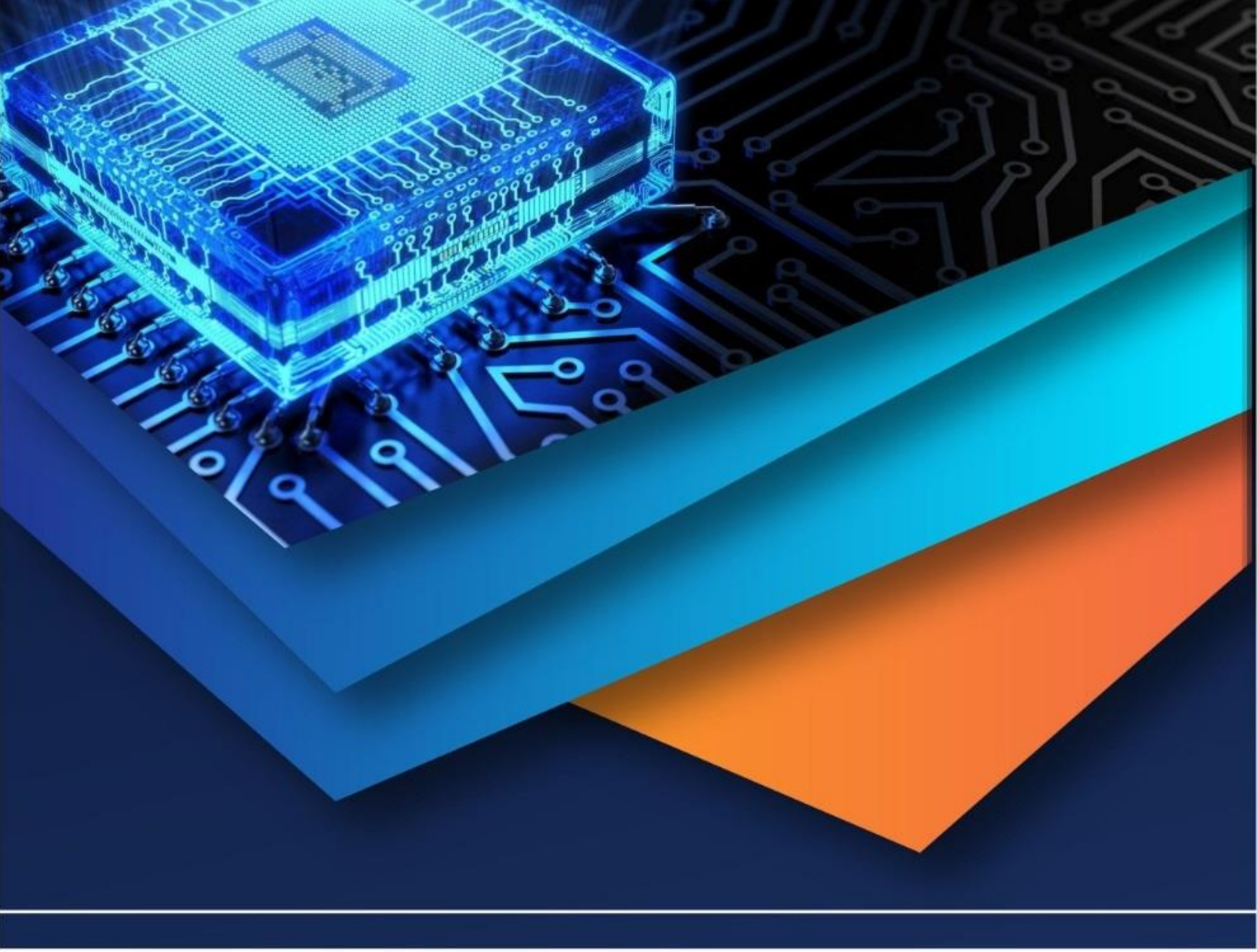
[8] Juba, B., & Le, H. S. (2019, July). Precision-recall versus accuracy and the role of large data sets. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 4039-4048).

[9] Persson, I., & Khojasteh, J. (2021). Python packages for exploratory factor analysis. Structural Equation Modeling: A Multidisciplinary Journal, 28(6), 983-988

[10] Gladence, L. M., Karthi, M., & Anu, V. M. (2015). A statistical comparison of logistic regression and different Bayes classification methods for machine learning. ARPN Journal of Engineering and Applied Sciences, 10(14), 5947-5953.

[11] Prihandi, I. KNN on Iris Data with Python Programming. vol, 2, 6-8.

[12] R.P, Prawin & R.P, Pranav & R, Swathi. (2023). Performance Evaluation and Comparative Analysis of Several Machine Learning Classification Techniques Using a Data-driven Approach in Predicting Renal Failure. International Journal for Research in Applied Science and Engineering Technology. 11. 3522-3530. 10.22214/ijraset.2023.54343.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY