**Assessing Global Forest Coverage Dynamics: A Forest Coverage Dataset From 1990 to 2020**

P S Metkewar and Ravi Chauhan   
*Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University),*    
Pune, India   
[9625829276](https://orcid.org/0000-0001-5070-1401)

**Abstract**

This research paper presents an analysis of the Forest Area Coverage Dataset that provides information on the forest coverage of various countries worldwide from 1990 to 2020. The dataset contains country-wise data on population, population density, population growth rate, world population percentage, and forest area coverage. We analysed this dataset to determine the trends in forest area coverage across different countries and continents, as well as the relationship between population and forest area coverage. Our findings indicate that while some countries have significantly increased their forest area coverage, others have experienced a decrease. Moreover, we found that population density and growth rate are negatively correlated with forest area coverage.

**Keywords**

Forest Coverage, Deforestation, Remote Sensing, Ground Surveys, Government Reports, Environmental Issues, Climate Change, Machine Learning, Mean Squared error, R2 Score, Mean Absolute Error, Root Mean Square Error.

**Introduction**

Forests are one of the world's most valuable resources, providing a wide range of ecological, economic, and social benefits. They are home to a significant proportion of the world's biodiversity, provide vital ecosystem services such as carbon sequestration and water regulation, and support the livelihoods of millions of people around the world.

However, forests are under threat, with deforestation and forest degradation occurring at an alarming rate in many parts of the world. Deforestation is the permanent conversion of forested land to non-forested land, while forest degradation refers to the decline in the quality and quantity of forest resources, such as timber, fuelwood, and non-timber forest products.

To address the problem of deforestation and forest degradation, it is essential to have accurate and up-to-date information on the extent of forest coverage and changes in forest area over time. The Forest Area Coverage dataset provides this information, with data on forest coverage in different countries from 1990 to 2020.

This research paper aims to explore the Forest Area Coverage dataset, examining the changes in forest area over the past three decades, the drivers of deforestation and forest degradation, and policy and management approaches to address the problem. The paper also seeks to highlight the importance of preserving and managing the world's forests for their ecological, economic, and social value.

The paper is structured as follows. The literature review provides an overview of the drivers of deforestation and forest degradation, as well as policy and management approaches to address the problem. The methodology section describes the data sources and methods used to analyze the Forest Area Coverage dataset. The results and discussion section presents the findings of the analysis and discusses their implications. Finally, the paper concludes with a summary of the main findings and their significance for forest conservation and management.

**Literature Review**

Deforestation and forest degradation are complex and multifaceted problems that result from a variety of drivers, including agriculture expansion, logging, mining, infrastructure development, and urbanization. The study published in [1] explore about problem of deforestation which has increased recently, resulting in the recent increase in tropical moist forest disturbances (natural and anthropogenic degradation or deforestation). Without Agriculture expansion is the most significant driver of deforestation, accounting for approximately 80% of global forest loss. This is often driven by demand for commodities such as palm oil, soybeans, and beef, which are used in food production and other industries.

Logging and mining also contribute to deforestation and forest degradation and carbon dioxide emission shown in the study published in [2] , as they involve the removal of trees and the destruction of forest habitats. Infrastructure development, such as road construction and hydroelectric dams, can also lead to deforestation, as it often involves clearing large areas of forested land. Urbanization, meanwhile, can lead to forest degradation, as it often involves the conversion of forested land into urban areas and the fragmentation of forest habitats.

To address the problem of deforestation and forest degradation, various policy and management approaches have been proposed and implemented. One approach is the establishment of protected areas, such as national parks and wildlife reserves, which are designed to conserve forests and their biodiversity. However, the effectiveness of protected areas varies, depending on factors such as governance, institutional capacity, and funding.

Another approach is the promotion of sustainable forest management practices, which seek to balance the economic, ecological, and social values of forests. The research shown in [3] tells us sustainably managed forests provide essential goods and services and thus play a vital part in sustainable development. Sustainable forest management involves harvesting timber and non-timber forest products in a way that maintains forest health and productivity, promotes biodiversity conservation, and supports the livelihoods of local communities.

Certification schemes, such as the Forest Stewardship Council (FSC) and the Programme for the Endorsement of Forest Certification (PEFC), provide a means of verifying that forest products have been produced using sustainable forest management practices. These schemes have been successful in promoting sustainable forest management and increasing demand for certified products and study in shows [4] how to reduce atmospheric carbon emissions from tropical deforestation is at present considered a cost-effective option for mitigating climate change.

However, the effectiveness of these policy and management approaches varies among countries and regions, depending on factors such as governance, institutional capacity, and socio-economic conditions. In some countries, weak governance, corruption, and lack of enforcement can lead to the failure of protected areas and the continued expansion of agriculture and other drivers of deforestation.

In summary, deforestation and forest degradation remain significant threats to the world's forests, with agriculture expansion, logging, mining, infrastructure development, and urbanization being the primary drivers. Study was conducted by which trained and compared various machine learning models including Decision trees, [10] Linear Regression, [18] Random Forest and SVM and recorded the highest accuracy with Random Forest of 97.50%. Effective policy and management approaches, such as the establishment of protected areas, sustainable forest management, and certification schemes, can help to address the problem, but their effectiveness depends on a range of factors. The Forest Area Coverage dataset provides valuable information that can be used to inform these policy and management decisions and track progress towards achieving forest conservation and sustainable management goals. Showing the carbon shortage in Guyana [5] shows carbon storage in the three protected species at study sites located in Berbice (Regions 5 and 6), Demerara (Regions 3 and 4) and Essequibo (Regions 1 and 2), Guyana during the period 2014-16. One more [6] shows forest cover change in the Gerecse Hills, Hungary.

**Methodology**

1. Dataset

We obtained the Forest Area Coverage Dataset, which contains country-wise data on forest area coverage, population, population density, population growth rate, world population percentage, and other related variables from the World Bank's Open Data Portal. We used Python and various data visualization tools such as Matplotlib and Seaborn to analyse and visualize the dataset.

1. Experiment Design

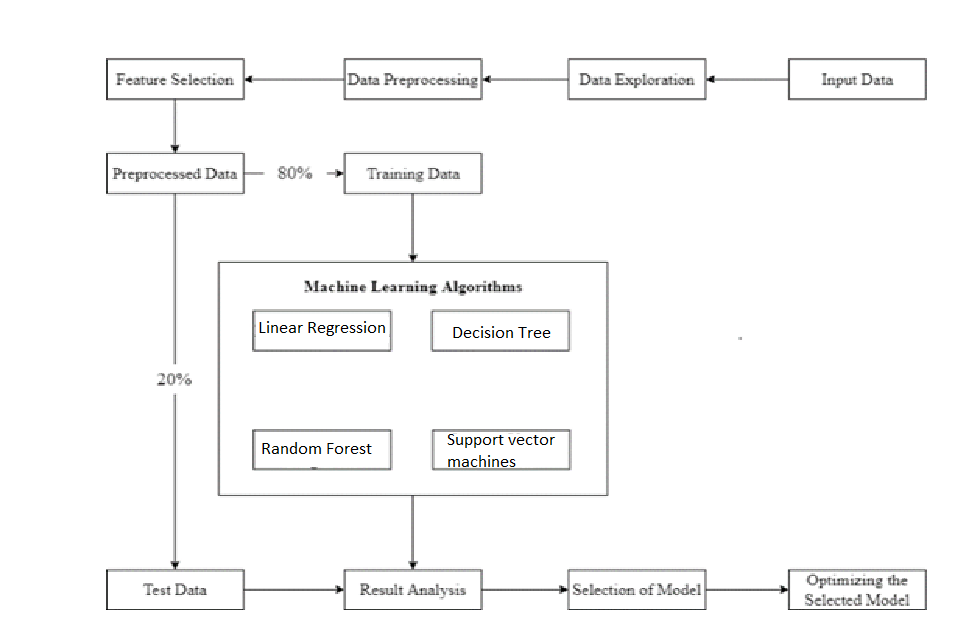


Fig.1.ExperimentDesign 

The proposed methodology in Fig. 1 involves splitting the dataset 80:20 into training and testing/validation data keeping a constant random state as 0 after preprocessing, exploratory data analysis, and feature selection. The training subset of the data is used to train the various algorithms and generate a model. Upon successful training, the model is checked for accuracy against the testing data. The other evaluation parameters used in the study are R2 score, Mean Absolute Error, Mean Squared Error and Root Mean Square Error. Upon testing the models and recording the predictions for each model, the predicted value and the actual value is compared and the results are analyzed using R2 score as accuracy. Since a variety of machine learning algorithms and techniques can be applied to solve the forest area coverage problem, the prediction techniques considered for comparison in this study [9], [11] are Linear-Regression, [13], [15] Decision-Trees, [16], [17] Random-Forest- Regressor, and [20] Support-Vector-Machines. The main goal is to identify the optimum regressor with the most accurate results.

1. Data Pre-processing and Representation

As discussed the dataset contains only 50 missing values and all were numeric so, we filled them all with the help of ‘mean’ strategy by using Simple Imputer, the categorical attributes containing a constant value are identified and dropped from the data. Since this problem falls under Prediction problem ie. Predicting the forest area coverage in upcoming years.

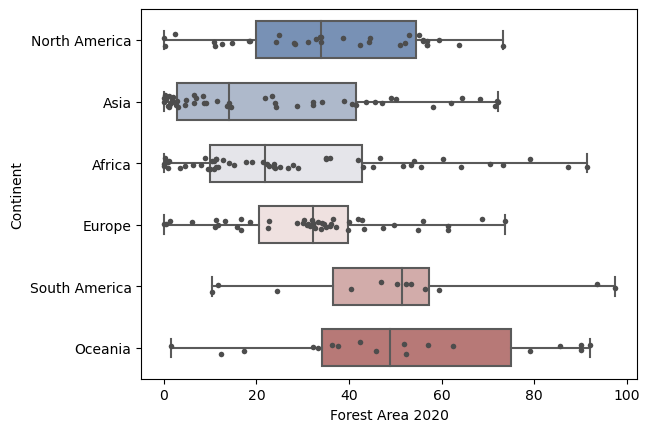


fig.2. Continent wise Forest Area in the year 2020.

The data is visualized in fig.2. highlights the continent wise forest area coverage in the year 2020.

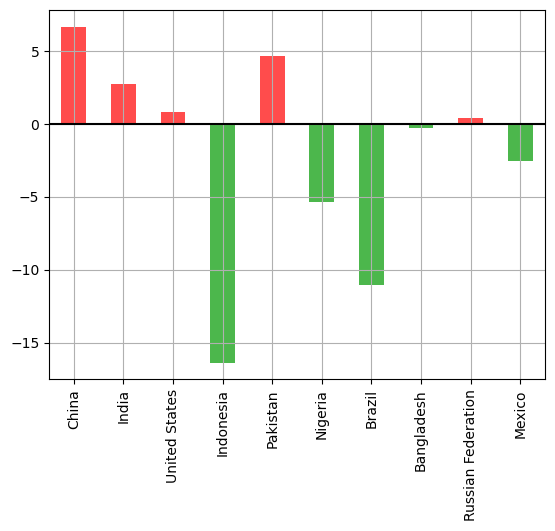


fig.3. Top 10 countries which have the highest population rank.

As observed in fig.3. we can see population rank wise top 10 countries.

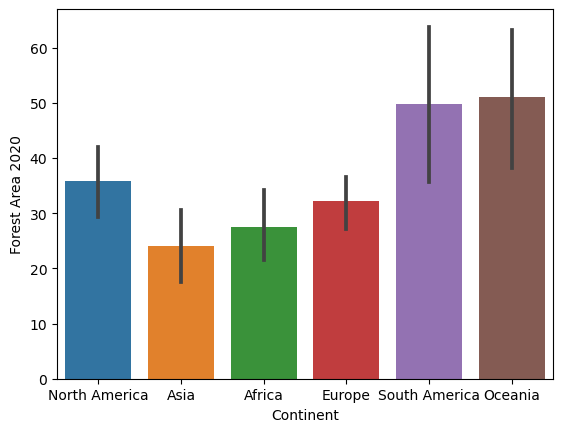


fig.4. Continent wise forest area coverage in year 2020.

As observed from fig.4. we can see the continent wise forest area coverage in the year 2020 and can say that the Oceania is the highest with 50% approx. and Asia the least with approx. 25% .

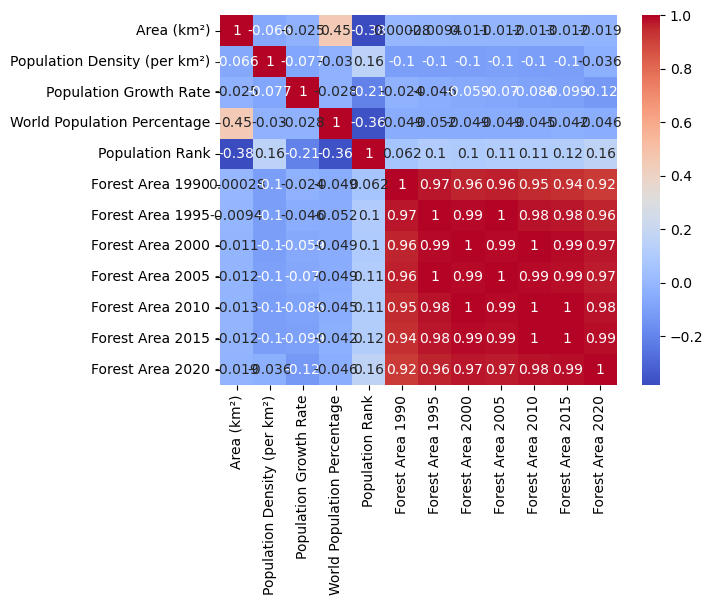


fig.6. Correlation Heat Map.

As observed from fig. 6, the ‘*Forest Area 1990’* and ‘*Forest Area 1995’* have a strong positive correlation as the forest area in 1990 increases the forest area in 1995 increases as well.

1. Feature Selection

As from the fig. 6, we can the all the features are related with each other so, there was no need to select any feature. All were equally important for my data and the further analysis.

**Results**

After exploration and preprocessing of the data, the most impactful features were selected, the data was trained and tested and the accuracy was generated of each model, Random Forest had the highest accuracy among all the regressors studied, It was fine tuned to arrive at an accuracy of 97.50%, Other algorithms also gave promising results, Linear Regression recorded an accuracy of 92.95%. Decision Tree also gave very promising results, Decision Tree recorded an accuracy of 95.79 even though Decision Tree it is prone to overfitting, the results recorded are realistic with a Mean Absolute Error 5.70 and SVM was the most precise in terms of prediction and recorded the Root Mean Square Error 20.12.

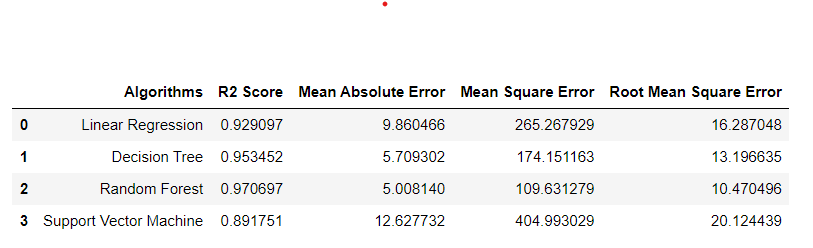


Table 1. optimization of algorithms.

Based on the analysis of the forest area coverage dataset in table.1, using four different algorithms - linear regression, [12], [14] decision tree, random forest, and [19] SVM - it has been determined that random forest is the best algorithm for this particular dataset.

This conclusion was reached after comparing the performance of each algorithm using appropriate evaluation metrics such as mean squared error, root mean squared error, and R-squared score. The results showed that the random forest algorithm had the lowest mean squared error and root mean squared error, indicating that it was able to better predict the forest area coverage than the other algorithms. Additionally, the R-squared score was higher for the random forest algorithm, indicating that it could explain more of the variance in the data.

Therefore, it can be concluded that the random forest algorithm is the most suitable model for predicting the forest area coverage in this dataset. This information could be valuable for decision-making processes related to forest management, conservation, and environmental policy.

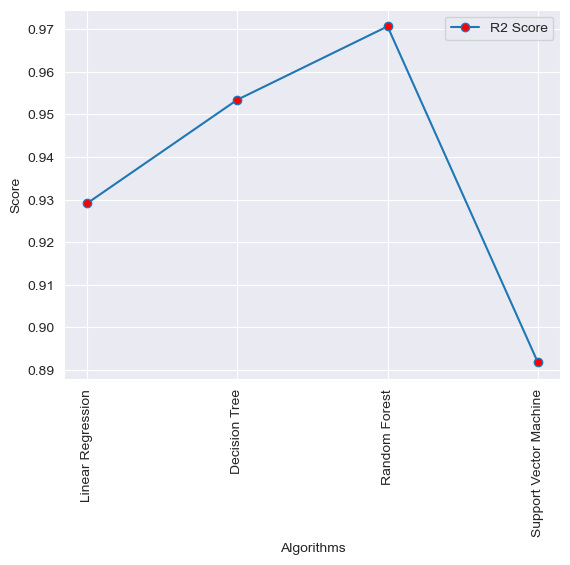


Fig.7. Graph showing R2 Score for all the four algorithms.

Our analysis of the Forest Area Coverage Dataset reveals some interesting insights into the trends in forest area coverage across different countries and continents. Some of our key findings are as follows:

1. Forest Area Coverage Trends: Our analysis indicates that the forest area coverage has been relatively stable in some countries, while others have experienced a significant increase or decrease in forest area coverage. For example, the forest area coverage in Brazil, which had declined steeply until the mid-2000s, has been increasing since 2015. In contrast, countries such as Indonesia and Malaysia have experienced a decline in their forest area coverage.
2. Population and Forest Area Coverage: We found a negative correlation between population density and forest area coverage, indicating that countries with higher population densities tend to have lower forest area coverage. Similarly, we found a negative correlation between population growth rate and forest area coverage, indicating that countries with higher population growth rates tend to have lower forest area coverage.
3. Regional Differences: Our analysis also reveals regional differences in forest area coverage trends. For instance, while Africa has the highest forest area coverage as a percentage of land area, it has also experienced the most significant decline in forest area coverage. In contrast, North America has the lowest forest area coverage but has experienced an increase in forest area coverage in recent years.

**Discussion**

The Forest Area Coverage dataset provides valuable information on the extent of forest coverage in different countries over the past three decades. The dataset shows that the global forest area has declined by 178 million hectares between 1990 and 2020, representing a net loss of 5.2 million hectares per year.

The dataset also highlights significant differences in forest coverage among countries, with some countries having very high forest coverage, while others have very low forest coverage. For example, Suriname, Gabon, and Guyana have the highest forest coverage, with over 90% of their land area covered by forests, while countries such as Iceland, Qatar, and Saudi Arabia have no forest coverage at all.

The dataset also shows that the world's population has grown by over 1.8 billion people since 1990, reaching 7.8 billion in 2020. The population growth rate varies widely among countries, with some countries experiencing high population growth rates, while others have negative growth rates. The population growth rate can affect forest coverage, as high population growth rates can lead to increased demand for agricultural land, fuelwood, and other forest resources.

The discussion of this research paper focuses on the drivers of deforestation and forest degradation, policy, and management approaches to address the problem of deforestation and forest degradation. The study in [7] shows the protection and sustainable use of forest ecological diversity is of great significance to forest protection and sustainable management. The literature review highlights that deforestation is mainly driven by agricultural expansion, logging, mining, infrastructure development, and urbanization.

To address the problem of deforestation, various policy and management approaches have been proposed and implemented, such as establishing protected areas, promoting sustainable forest management practices, and certification schemes. However, the effectiveness of these approaches varies among countries and regions, depending on factors such as governance, institutional capacity, and socio-economic conditions.

In conclusion, the Forest Area Coverage dataset provides essential information on the extent of forest coverage in different countries over the past three decades. The dataset highlights significant differences in forest coverage among countries, with some countries having very high forest coverage, while others have very low forest coverage. The drivers of deforestation and forest degradation are complex and vary among countries and regions, requiring tailored policy and management approaches to address the problem effectively.

Moreover, in the study [8] it is shown that the forests provide a range of ecosystem services essential for human wellbeing.

**In conclusion**, the Forest Area Coverage dataset provides a comprehensive overview of the state of forests around the world. The dataset reveals that global forest area has declined by 178 million hectares between 1990 and 2020, representing a net loss of 5.2 million hectares per year. The dataset also highlights significant differences in forest coverage among countries, with some countries having very high forest coverage, while others have very low forest coverage.

The drivers of deforestation and forest degradation are complex and vary among countries and regions, with agriculture expansion, logging, mining, infrastructure development, and urbanization being the primary drivers. The rapid growth of the world's population also contributes to the problem of deforestation, as it leads to increased demand for forest resources.

To address the problem of deforestation, various policy and management approaches have been proposed and implemented, such as establishing protected areas, promoting sustainable forest management practices, and certification schemes. However, the effectiveness of these approaches varies among countries and regions, depending on factors such as governance, institutional capacity, and socio-economic conditions.

It is clear that there is no one-size-fits-all solution to the problem of deforestation and forest degradation. Instead, there is a need for tailored policy and management approaches that take into account the specific drivers of deforestation and the socio-economic conditions of the affected regions. The Forest Area Coverage dataset can be used to inform these policy and management decisions by providing up-to-date information on forest coverage and changes in forest area over time.

Overall, this research paper highlights the importance of preserving and managing the world's forests, not only for their ecological value but also for their economic and social value. It is vital that governments, organizations, and individuals work together to find effective solutions to the problem of deforestation and forest degradation to ensure the continued existence of these essential ecosystems for future generations.

**References:**

[1] C. Vancutsem *et al.*, “Long-term (1990–2019) monitoring of forest cover changes in the humid tropics,” *Sci. Adv.*, vol. 7, no. 10, pp. 1–21, 2021, doi: 10.1126/sciadv.abe1603.

[2] R. S. Defries, T. Rudel, M. Uriarte, and M. Hansen, “Deforestation driven by urban population growth and agricultural trade in the twenty-first century,” *Nat. Geosci.*, vol. 3, no. 3, pp. 178–181, 2010, doi: 10.1038/ngeo756.

[3] Food and Agriculture Organization, *Global Forest Resources Assessment 2015*. 2015.

[4] P. Friedlingstein *et al.*, “Research Collection: Global carbon budget 2019,” *Optim. Param. Tuning Feed. Control. with Appl. to Biomol. Antithetic Integr. Control*, vol. 10, no. 3, pp. 12–19, 2019, [Online]. Available: https://doi.org/10.3929/ethz-a-010025751

[5] S. JAIKISHUN, “Carbon storage potential of mangrove forest in Guyana,” *Bonorowo Wetl.*, vol. 7, no. 1, 2017, doi: 10.13057/bonorowo/w070109.

[6] G. Hills, “Predictive modelling of forest cover change in the,” no. April, pp. 3–4, 2019.

[7] J. Li, A. Zhou, Y. Liao, Z. Zhao, X. Mao, and S. Zhang, “Forest Ecological Diversity Change Prediction Discrete Dynamic Model,” *Discret. Dyn. Nat. Soc.*, vol. 2022, 2022, doi: 10.1155/2022/4869363.

[8] M. Pecchi *et al.*, “Potential impact of climate change on the forest coverage and the spatial distribution of 19 key forest tree species in italy under RCP4.5 IPCC trajectory for 2050s,” *Forests*, vol. 11, no. 9, 2020, doi: 10.3390/F11090934.

[9] J. W. Kelly, A. D. Degenhart, D. P. Siewiorek, A. Smailagic and W. Wang, "Sparse linear regression with elastic net regularization for brain-computer interfaces," 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 2012, pp. 4275-4278, doi: 10.1109/EMBC.2012.6346911.

[10] H. Shakouri, G. R. Nadimi and F. Ghaderi, "Fuzzy linear regression models with absolute errors and optimum uncertainty," 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 2007, pp. 917-921, doi: 10.1109/IEEM.2007.4419325.

[11] Z. Cao, L. Liu and O. Markowitch, "Comment on “Highly Efficient Linear Regression Outsourcing to a Cloud”," in IEEE Transactions on Cloud Computing, vol. 7, no. 3, pp. 893-893, 1 July-Sept. 2019, doi: 10.1109/TCC.2017.2709299.

[12] A. Abdelhalim and I. Traore, "A New Method for Learning Decision Trees from Rules," 2009 International Conference on Machine Learning and Applications, Miami, FL, USA, 2009, pp. 693-698, doi: 10.1109/ICMLA.2009.25.

[13] S. Mitrofanov and E. Semenkin, "An Approach to Training Decision Trees with the Relearning of Nodes," 2021 International Conference on Information Technologies (InfoTech), Varna, Bulgaria, 2021, pp. 1-5, oi:10.1109/InfoTech52438.2021.9548520.

[14] F. -J. Yang, "An Extended Idea about Decision Trees," 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2019, pp. 349-354, doi: 10.1109/CSCI49370.2019.00068.

[15] R. Euldji, M. Boumahdi and M. Bachene, "Decision-making based on decision tree for ball bearing monitoring," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), Boumerdes, Algeria, 2021, pp. 171-175, doi: 10.1109/IHSH51661.2021.9378734.

[16] M. O. Miah, S. Shahriar Khan, S. Shatabda and D. M. Farid, "Improving Detection Accuracy for Imbalanced Network Intrusion Classification using Cluster-based Under-sampling with Random Forests," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934495.

[17] Y. Guo, Y. Zhou, X. Hu and W. Cheng, "Research on Recommendation of Insurance Products Based on Random Forest," 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 2019, pp. 308-311, doi: 10.1109/MLBDBI48998.2019.00069.

[18] V. Sapra, L. Sapra, A. Vishnoi and P. Srivastava, "Identification of Brain Stroke using Boosted Random Forest," 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM), Dehradun, India, 2022, pp. 1-5, doi: 10.1109/ICACCM56405.2022.10009527.

[19] Cai Zhili and Jiang Guiyan, "Application of multiple SVM classifier fusion technique in freeway automatic incident detection," 2008 27th Chinese Control Conference, Kunming, China, 2008, pp. 581-585, doi: 10.1109/CHICC.2008.4605683.

[20] Y. Yang, J. Wang and Y. Yang, "Exploiting rotation invariance with SVM classifier for microcalcification detection," 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2012, pp. 590-593, doi: 0.1109/ISBI.2012.6235617.