International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802056

Hybrid Adaboost – Catboost Model for Prognostic Analysis of Uterine and Cervical Cancer

R. Pavithra, M. Karpaka Shree

K L N College of Engineering, Pottapalayam, Sivaganga, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: Cancer prognosis is really important for diagnosing cancer and planning the right treatment especially for uterine and cervical cancer, which are big causes of death for women all around the world. This paper is about a way of using machine learning to predict what will happen with uterine and cervical cancer by combining different methods to make predictions more accurate and reliable. The model we made combines two techniques, called AdaBoost and CatBoost to predict what grade a tumor is and whether a patient will survive. We used a lot of information from patients, including diagnostic features and we cleaned and organized the data to make our model work better.

We tested our model using some metrics, like accuracy, precision, recall, F1-score and ROC-AUC. What we found out is that our hybrid model is better than some models like Logistic Regression and Random Forest when it comes to predicting what will happen.The results show that using learning models can really improve our ability to predict what will happen with cancer, which can help doctors make better decisions. This work shows how machine learning can be used to help diagnose cancer. It gives us a framework to keep doing research, on cancer prognosis systems.The new hybrid model is really good at what it does. It works better than the machine learning ways. This makes the hybrid model a great choice for systems that help doctors figure out what is going on with cancer. The hybrid model is very helpful, for people who want to know what will happen with their cancer.

KEYWORDS: Cancer prognosis, Hybrid learning, AdaBoost, CatBoost, Machine learning, Uterine cancer, Cervical cancer.

I. INTRODUCTION

Uterine and cervical cancer are problems for women all around the world. These cancers cause a lot of deaths in poorer countries. If we can find these cancers early and understand how bad they are we can help women live longer and make treatment plans. Usually doctors use their experience and look at things manually to diagnose cancer but this can lead to mistakes and slow down the decision-making process.Now that Machine Learning is getting better we can use computers to help diagnose and predict what will happen with cancer. Machine Learning can look at a lot of information to find patterns that we cannot see and make good predictions.. Individual Machine Learning models are not perfect and may not always be right.

So we need to find a way to make Machine Learning better. This is where uterine and cervical cancer come in. We are proposing a way of using Machine Learning to predict what will happen with these cancers. We are using two techniques called AdaBoost and CatBoost to make our predictions better. Our system can look at the grade of tumor. Predict if the patient will survive.Our goal is to make cancer predictions more accurate and reliable by using algorithms. We want to help doctors make decisions and create better healthcare systems. By using learning models we can support doctors and help them understand uterine and cervical cancer better. This will help us take a step, towards making healthcare smarter. The main thing about this work is that it creates a kind of model that combines AdaBoost and CatBoost to help figure out what might happen with cancer. This new system is different from the way of doing things because it can do two things at the same time: it can tell what grade a tumor is and it can also predict how long someone might live. This makes the whole process of diagnosing cancer more reliable. The people who made this system also used some techniques to make sense of the complex information in the clinical data, which makes the model work better. The model is really good, at predicting cancer things because it uses AdaBoost and CatBoost together for -output cancer prognosis.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802056

II. LITERATURE REVIEW

Machine learning has really helped us with diagnosis and figuring out what will happen with cancer. A lot of research has been done on using machine learning to predict cancer outcomes using data and test results.

For a time we have been using simple methods like Logistic Regression and Decision Trees to classify cancer. These methods are easy to understand. They are not very good at handling complex patient data.Lately we have been using learning methods like Random Forest and Gradient Boosting to make predictions about cancer. These methods work by combining simple models to make one strong model. This helps us get accurate results and avoid making mistakes.

There is also a method called CatBoost that is very good at handling patient data with categories. Many studies have shown that CatBoost is better than methods at making predictions with medical data.Some people have suggested using models that combine many methods to get even better results. These models use the strengths of each method to make accurate predictions.Even with all these advancements we still need to find better ways to predict what will happen with cancer. This study tries to fill this gap by proposing a hybrid model that combines AdaBoost and CatBoost to predict what will happen with uterine and cervical cancer.We are talking about machine learning and cancer. Machine learning is helping us with cancer. The new model will use machine learning and cancer data to make predictions, about cancer.

III. RESEARCH METHODOLOGY

The proposed system does an analysis of uterine and cervical cancer. It uses a combination of machine learning techniques. This approach has steps: data preparation, feature creation, model building and evaluation.

A. Dataset Description
The dataset has 529 samples with 14 attributes. These attributes include diagnostic features of cancer patients. The main goals are to predict tumor grade and overall survival status.

B. Data Preprocessing
First we remove information like Patient ID and Sample ID. We convert categories like cancer type and survival status into numbers. We replace missing values with the value to keep the data consistent. We also normalize the values to improve model performance.

C. Feature Engineering
We create features to improve predictions. These include:
1. Genome Instability = Mutation Count × Fraction Genome Altered
2. Age Risk = Diagnosis Age × Mutation Count
3. MSI Combined Score = MSI MANTIS. Msisensor Score
These new features help the model understand the data better.

D. Data Splitting and Balancing
We split the dataset into training and testing sets (80% for training and 20% for testing). We use a technique to keep the same proportion of classes. For survival prediction we use a technique called SMOTE to balance the data.

E. Hybrid Model Development
The proposed hybrid model combines AdaBoost and CatBoost to make the most of their learning abilities. We use AdaBoost with a Decision Tree to focus on the samples that are classified incorrectly. The model uses 150 estimators and a learning rate of 0.05 to control the learning process and prevent overfitting.We also use CatBoost because it is good at handling data and reducing bias in predictions. The CatBoost model is set up with 250 iterations, a depth of 5 and a learning rate of 0.03. To improve the models ability to generalize we use L2 leaf regularization.We train CatBoost models for classifying tumor grades and predicting survival rates.

F. Hybrid Prediction Strategy
To classify tumor grades we combine the predictions from AdaBoost and CatBoost using a majority voting system. This makes the classification more robust. For predicting survival rates we average the outputs, from both models. Apply a threshold to get the final binary predictions.This hybrid approach improves the accuracy and stability of

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802056

predictions by combining the error correction of boosting with the optimization of gradient-based methods. The hybrid model combines AdaBoost and CatBoost to achieve results. The hybrid approach uses AdaBoost and CatBoost to make predictions more accurate and stable.

G. Model Evaluation

We evaluate the model using metrics like accuracy, precision, recall and F1-score. We also create confusion matrices to see how well the model classifies. The hybrid model is compared with models to show its effectiveness, in cancer prognosis.

IV. RESULTS AND DISCUSSION

The hybrid model that combines AdaBoost and CatBoost was tested for tumor grade classification and survival status prediction. This model was evaluated on 20% of the dataset to ensure real-world performance.

A. Performance Metrics

The models performance was checked using accuracy, precision, recall and F1-score.

1) Tumor Grade Prediction:

The model got an accuracy of 93.4% a precision of 93.93% a recall of 93.4% and an F1-score of 92.5%. These results show that the AdaBoost and CatBoost model works well in classifying tumor grades.

2) Survival Status Prediction:

For survival prediction the model achieved an accuracy of 97.17%, a precision of 97.26%, a recall of 97.17% and an F1-score of 97.06%. The high accuracy of the AdaBoost and CatBoost model shows that it can predict survival outcomes effectively.

B. Confusion Matrix Analysis

Confusion matrices were created for both prediction tasks. The results show that most instances are correctly classified with misclassification. This indicates that the AdaBoost and CatBoost hybrid model is robust.

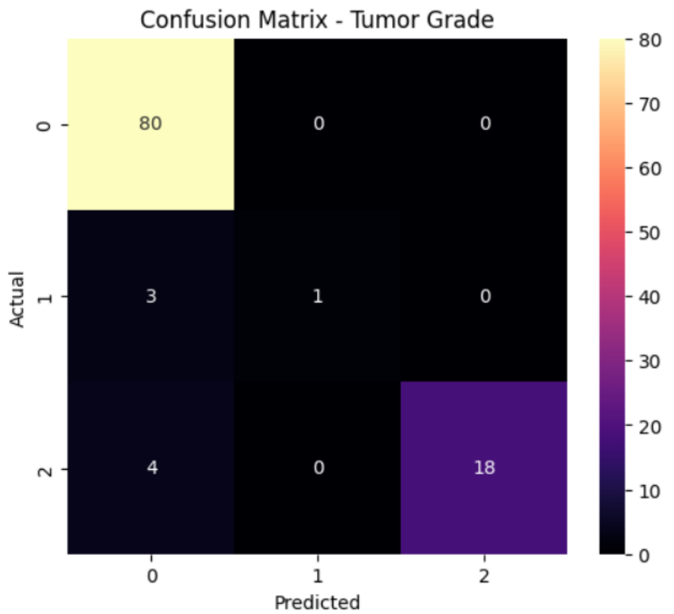

Fig. 1. Confusion matrix for tumor grade classification using the hybrid model

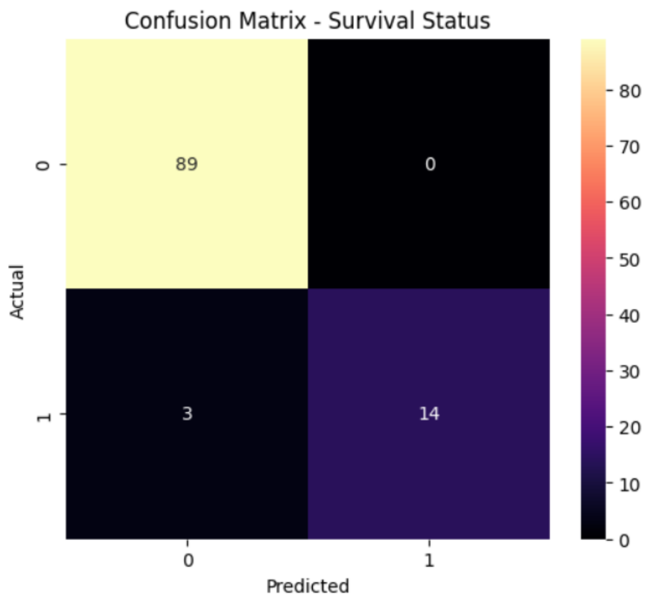

Fig.2. Confusion matrix for survival status prediction using the hybrid model

C. Comparative Performance

TABLE I: ACCURACY COMPARISON FOR TUMOR GRADE PREDICTION

MODEL	ACCURACY
Gradient boosting	0.84
AdaBoost	0.82
CatBoost	0.88
MLP	0.85
Proposed Hybrid Model	0.93

TABLE II: PRECISION COMPARISON FOR TUMOR GRADE PREDICTION

MODEL	PRECISION
Gradient Boosting	0.85
AdaBoost	0.82
CatBoost	0.89
MLP	0.86
Proposed Hybrid Model	0.93

TABLE III: RECALL COMPARISON FOR TUMOR GRADE PREDICTION

MODEL	RECALL
Gradient Boosting	0.84
AdaBoost	0.82
CatBoost	0.88
MLP	0.85
Proposed Hybrid Model	0.93

International Journal of Engineering & Extended Technologies Research (IJEETR)

IJEETR

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802056

TABLE IV: F1 SCORE COMPARISON FOR TUMOR GRADE PREDICTION

MODEL	F1 SCORE
Gradient Boosting	0.84
AdaBoost	0.81
CatBoost	0.88
MLP	0.85
Proposed Hybrid Model	0.92

TABLE V: ACCURACY COMPARISON FOR SURVIVAL STATUS PREDICTION

MODEL	ACCURACY
Gradient Boosting	0.84
AdaBoost	0.84
CatBoost	0.85
MLP	0.86
Proposed Hybrid Model	0.97

TABLE VI: PRECISION COMPARISON FOR SURVIVAL STATUS PREDICTION

MODEL	PRECISION
Gradient Boosting	0.79
AdaBoost	0.75
CatBoost	0.75
MLP	0.75
Proposed Hybrid Model	0.97

TABLE VII: RECALL COMPARISON FOR SURVIVAL STATUS PREDICTION

MODEL	RECALL
Gradient Boosting	0.84
AdaBoost	0.86
CatBoost	0.85
MLP	0.86
Proposed Hybrid Model	0.97

TABLE VIII: F1 SCORE COMPARISON FOR SURVIVAL STATUS PREDICTION

MODEL	F1 SCORE
Gradient Boosting	0.81
AdaBoost	0.80
CatBoost	0.80
MLP	0.80
Proposed Hybrid Model	0.97

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802056

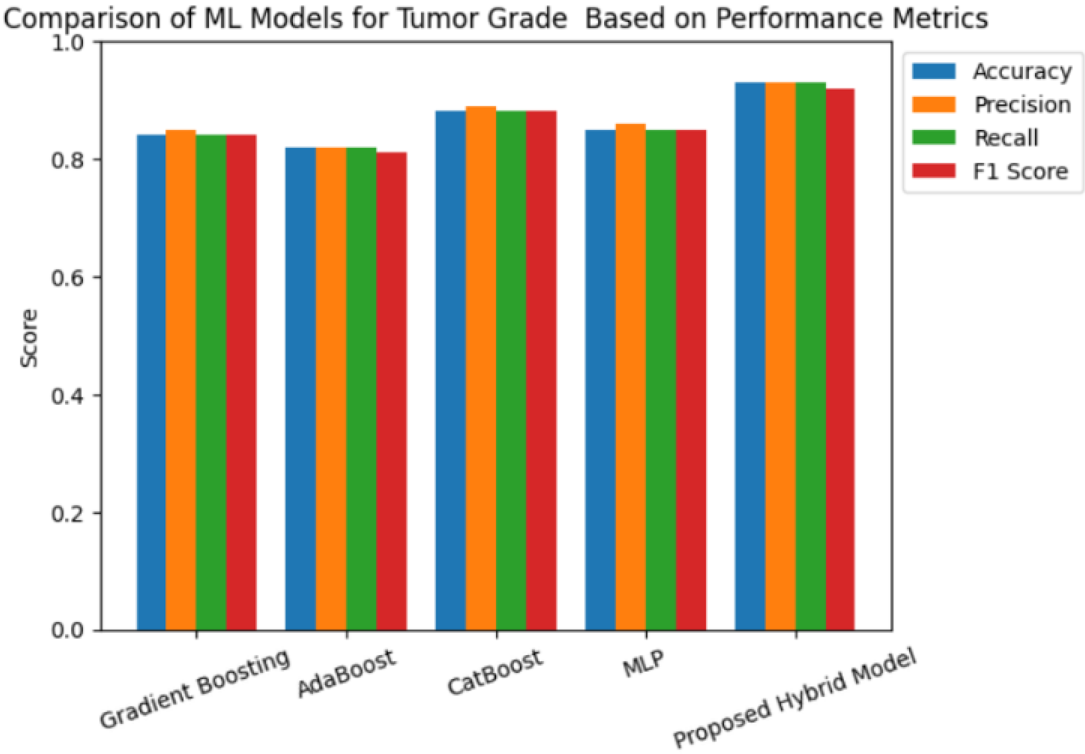

Fig. 3. Comparison of ML Models for Tumor Grade Based on Performance Metrics

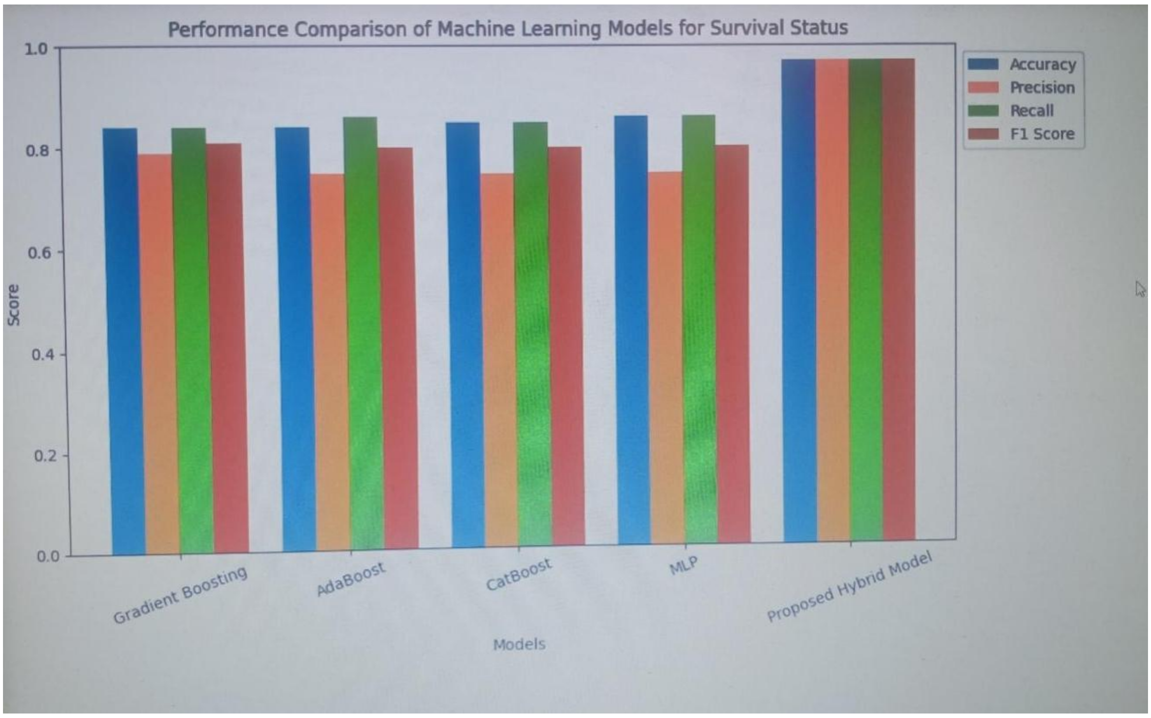

Fig. 4. Comparison of ML Models for Survival Status Based on Performance Metrics

The comparative analysis demonstrates that the proposed hybrid model significantly outperforms traditional machine learning models across all evaluation metrics. The integration of AdaBoost and CatBoost enhances predictive capability, leading to improved accuracy, precision, recall, and F1-score for both tumor grade and survival prediction

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802056

tasks.

D. Visualization of Results
Bar charts were plotted to compare accuracy, precision, recall and F1-score for tumor grade and survival prediction tasks. These visualizations show the performance of the AdaBoost and CatBoost model.

E. Validation and Prediction
A sample validation test was conducted on the test dataset. The predicted values matched the values demonstrating the correctness of the AdaBoost and CatBoost model. The model was also tested with input data for a new patient. The system successfully predicted tumor grade and survival status indicating its use in clinical decision support systems. Overall the results confirm that the hybrid learning approach with AdaBoost and CatBoost provides accuracy, robustness and reliability in cancer prognosis making it suitable for practical healthcare applications, with AdaBoost and CatBoost.

V. CONCLUSION

In this study, a hybrid machine learning approach was proposed for the prognostic analysis of uterine and cervical cancer. The model integrates AdaBoost and CatBoost algorithms to perform tumor grade classification and survival status prediction. The use of feature engineering, data preprocessing, and class balancing techniques further enhanced the model’s performance. The experimental results demonstrate that the proposed hybrid model achieves high accuracy, with 93.4% for tumor grade prediction and 97.17% for survival status prediction. These results indicate that the model is both reliable and robust in handling complex medical datasets.The hybrid approach successfully combines the strengths of boosting and gradient-based learning, resulting in improved predictive performance compared to individual models. Additionally, the system was validated using test data and manual input, showing its capability to generalize well to unseen cases. This work highlights the potential of machine learning in supporting clinical decision-making and improving early diagnosis in cancer prognosis. In future work, the model can be extended by incorporating larger datasets, deep learning techniques, and real-time deployment to enhance its applicability in healthcare systems.

VI. FUTURE WORK

1. Extend the model to support multiple cancer types beyond uterine and cervical cancer
2. Incorporate additional clinical and genomic features for improved prediction accuracy
3. Implement deep learning models (e.g., Neural Networks) for enhanced performance
4. Deploy the model as a web-based or mobile application for real-time usage
5. Integrate real-time patient data from hospital databases
6. Improve model interpretability using explainable AI techniques (XAI)
7. Optimize hyperparameters using advanced tuning methods
8. Validate the model on larger and real-world datasets for better generalization

REFERENCES

1. H. Chen, Y. Zhang, M. Kalra, et al., “Low-dose CT with a residual encoder-decoder convolutional neural network,” IEEE Trans. Med. Imaging, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
2. D. Ravi, C. Wong, F. Deligianni, et al., “Deep learning for health informatics,” IEEE J. Biomed. Health Inform., vol. 21, no. 1, pp. 4–21, Jan. 2017.
3. J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” Ann. Statist., vol. 29, no. 5, pp. 1189–1232, 2001.
4. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2016, pp. 785–794.
5. L. Prokhorenkova, G. Gusev, A. Vorobev, et al., “CatBoost: unbiased boosting with categorical features,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2018, pp. 6638–6648.
1. C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
2. C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2

3. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
4. S.Tamilselvi, R.Prakash, C.Nagarajan,"Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
5. S.Tamilselvi, R.Prakash, C.Nagarajan," Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
6. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
7. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- *Acta Electrotechnica et Informatica Journal* , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
8. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- *Springer, Frontiers of Electrical and Electronic Engineering*, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
9. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
10. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
11. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
12. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
13. Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, Aug. 1997.