



AI-Based Prediction of Beneficiary Eligibility for Government Welfare Schemes

Leelavathi V, Pavithra S, Dr. Sanju Rajan

Department of Computer Science and Engineering, Vels Institute of Science, Technology & Advanced Studies
(VISTAS), Chennai, India

Department of Computer Science and Engineering, Vels Institute of Science, Technology & Advanced Studies
(VISTAS), Chennai, India

Department of Computer Science and Engineering, Vels Institute of Science, Technology & Advanced Studies
(VISTAS), Chennai, India

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

ABSTRACT: Government welfare programs are of great use in making people's lives better, especially in rural areas. However, it is not always easy to find the right people who need government welfare programs, especially since the process requires manual intervention, incomplete information, and lack of knowledge among people. The project proposed here is related to finding out whether people are eligible for government welfare programs or not by using an AI-based system. The system will use machine learning to find out whether a person is eligible for a specific program by asking for attributes such as income level, family background, etc. This model will help find out whether a person is eligible for a specific program by processing all the information automatically. Machine learning will help eliminate errors made by people, provide accuracy, and ensure fair distribution of government benefits among people. The system proposed here is simple and easy to use, so it will also help people in rural areas who do not know much about technology.

KEYWORDS: Artificial Intelligence, Machine Learning, Government Welfare Schemes, Eligibility Prediction, Rural Development

I. INTRODUCTION

The main aim of the government schemes is to provide assistance to poor and financially weak individuals. These schemes provide assistance in the form of money, health, education, and housing. In India, these schemes are implemented for the rural and disadvantaged population. However, it is challenging to determine the actual beneficiaries of these schemes. At present, the information is verified manually, and it is challenging to check information such as income, occupation, and background. Thus, the actual beneficiaries do not receive the schemes, and the wrong individuals receive them. To solve this problem, technologies such as artificial intelligence and machine learning can be implemented. These technologies have already been proposed in recent research [2]. In this project, a system is developed for selecting the right individuals for the schemes. This helps in increasing accuracy, reducing human errors, and ensuring fair distribution of benefits.

A. Problem Statement

The welfare programs offered by the government help the poor and those suffering from various social problems. However, it is still very challenging for the government to select the correct individuals who deserve these welfare programs. Currently, the government checks the documents and applications manually. This process is very tedious and time-consuming for the government. There is a possibility of errors if the data is incorrect or not sufficient. Unfortunately, the deserving individuals do not get the required help because they are not aware of the welfare programs, the data may be incorrect, or the process may take too long. There is a possibility that the individuals who do not deserve these welfare programs may get them due to errors caused by humans. We need an intelligent system that can automatically determine whether an individual deserves a particular welfare program offered by the government or not based on their socioeconomic status. This process can be made precise, less tedious, transparent, and effective for the government while selecting individuals for welfare programs.



B. Motivation

The motivation for doing this project arises from the challenges faced in identifying the correct beneficiaries for government welfare schemes. The difficulties in determining who should receive government welfare programs are the driving force behind this project. Due to lack of knowledge, complicated processes, and mistakes in manual verification, many people, especially those living in rural areas, are frequently not receiving the benefits. However, some ineligible people might get benefits because of misinformation or human errors. This circumstance emphasizes the need for a beneficiary identification system that is more dependable and effective. It will be more feasible to evaluate socioeconomic data and make more precise eligibility predictions by using technologies like artificial intelligence (AI) and machine learning (ML). Our project's primary goal is to create a user-friendly and intelligent system that helps to determine the eligible citizens, minimize human error, and guarantee fair distribution among the people for the government welfare benefit schemes.

C. Objectives

The primary objective of this project is to build an intelligent system for predicting welfare scheme eligibility through various machine learning tools such as XG Boost and NLP. This will help process structured data such as income levels, age, occupation, and land details. In addition, it will process unstructured data such as guidelines for welfare schemes and details regarding applicants. The project also aims at reducing the workload and errors associated with humans through an automated process. It will make the selection process transparent and efficient. In addition, the project aims at building a simple and user-friendly interface so that it can be used for rural areas for the proper distribution of welfare benefits.

II. EXISTING SYSTEMS

The current system for selecting beneficiaries of government welfare schemes mainly depends on manual processes. Applicants are required to submit documents like ID proof, income certificates, and other supporting papers, which are then verified by officials to decide eligibility. This process takes a lot of time and effort, especially when there are many applicants. Due to this, errors can occur where eligible people may be rejected because of missing documents, and ineligible people may get selected. Also, the system mostly uses only basic structured data and does not consider large or unstructured data sources. Overall, the existing method is slow, less transparent, and sometimes inaccurate, which creates a need for a better and more efficient system that can handle large data and improve the selection process.

A. Research Papers and Studies

Many researchers have studied the use of artificial intelligence and machine learning in identifying beneficiaries for welfare schemes. One study proposed a machine learning-based system to automatically determine whether a person is eligible for social welfare programs, and it showed better accuracy compared to traditional methods. Another study tested different ML algorithms such as Naïve Bayes, Random Forest, Support Vector Machine, and K-Nearest Neighbor to classify households based on their socio-economic status and found that these methods can help governments in selecting the right beneficiaries more effectively. Some research has also focused on detecting false income claims using machine learning, which helps ensure that only eligible people receive benefits. In addition, studies using the XGBoost algorithm have shown that it can handle complex data relationships and provide more accurate predictions compared to traditional approaches. Overall, these studies indicate that AI and ML can play an important role in improving the efficiency and accuracy of welfare beneficiary selection systems.

B. Identified Gaps

Many studies use machine learning to choose who gets help from welfare programs. But these methods have problems. They mostly use organized data about people's economic status and ignore unstructured text data like program details or what applicants say. This is a big gap. Also, some models struggle with complex data that has many variables, which can lower accuracy. Many systems are not easy to use, especially for people in rural areas who may not know much about technology. Plus, many steps still need people to do them by hand, which takes time and can lead to mistakes. So, we need a better system that uses advanced methods like XGBoost and Natural Language Processing (NLP). This system can look at both organized and unstructured data, making it more accurate, clear, and efficient in choosing the right people for help.



III. EXISTING SYSTEM ANALYSIS

Currently, the process of selecting beneficiaries for government welfare schemes is mostly manual. Applicants submit documents like income proof and ID, which are checked by officials to decide eligibility. This process is time-consuming and requires a lot of effort, especially when there are many applications. Due to manual handling, errors can occur, such as missing details or incorrect verification. The system also struggles to manage large numbers of applications efficiently, causing delays in providing benefits. Moreover, it mainly depends on basic structured data and does not use advanced analysis to improve decision-making. Because of these limitations, the existing system is slow, less accurate, and not scalable. This creates a need for an automated and intelligent system that can analyze data effectively and select beneficiaries in a faster and more reliable way.

A. Limitations

At present, the selection of beneficiaries for government welfare schemes is mainly done through a manual process. Applicants submit documents such as income certificates and ID proofs, which are then verified by government officials to determine eligibility. This process takes a lot of time and effort, especially when there are many applications to review. Since the work is done manually, there is a high chance of errors, such as missing information or incorrect verification. The system also struggles to handle a large number of applications, leading to delays in providing benefits to deserving people. In addition, the current method relies only on basic structured data and does not make use of advanced techniques to analyze patterns or improve decision-making. Overall, the existing system is slow, less accurate, and not suitable for handling large-scale data. Therefore, there is a need for an automated and intelligent system that can process data efficiently and identify the right beneficiaries in a faster and more reliable way.

B. Proposed Improvements

This project improves the existing system by introducing an automated method to predict eligibility for government welfare schemes. Instead of manual checking, it uses the XGBoost algorithm to analyze socio-economic data like income, age, occupation, and family details. It also applies Natural Language Processing (NLP) to understand text data such as scheme descriptions and applicant information. The system reduces manual effort and errors while improving accuracy. It is designed to be simple and user-friendly, making it suitable even for rural users. The whole setup is designed to be straightforward and user-friendly, which means people in rural areas—especially those who aren't tech-savvy—can easily use it. By combining machine learning with data analysis, this system should make welfare distribution more accurate, transparent, and fair.

TABLE I. COMPARISON OF EXISTING AND PROPOSED SYSTEM

Aspect	Existing System	Proposed System
Process	Manual verification	Automated using ML
Data Handling	Structured data only	Structured + unstructured data
Technology	Traditional methods	XGBoost and NLP
Accuracy	Lower (human errors)	Higher (data-driven prediction)
Time Efficiency	Time-consuming	Faster processing
Bias	Possible human bias	Reduced bias
Scalability	Limited	Highly scalable
Accessibility	Less user-friendly	Simple and user-friendly



IV. PROPOSED SYSTEM

This system uses a smart approach to predict eligibility for government welfare schemes using machine learning. It analyzes socio-economic details such as income, age, occupation, and family size provided by applicants. The XGBoost algorithm studies this data to identify patterns and classify applicants as eligible or not eligible. In addition, the system uses Natural Language Processing (NLP) to understand unstructured text like scheme descriptions and applicant inputs. Combining ML and NLP improves the accuracy of predictions. The main aim is to reduce manual work and errors while making the process more transparent and fair. The system is also designed to be simple and easy to use, even for people in rural areas, ensuring that benefits reach the right individuals efficiently.

A. System Architecture

This system is designed to identify eligible beneficiaries for government welfare schemes in an efficient way. It first collects applicant details such as age, income, occupation, education, family size, and land ownership to form a structured dataset. It also considers unstructured data like scheme descriptions, eligibility rules, and applicant inputs. Next, the data is preprocessed by removing errors, handling missing values, and converting it into a suitable format. NLP techniques are used to extract important information from text data. The processed data is then given to the XGBoost model, which analyzes patterns and predicts eligibility. Finally, the system displays the result clearly to the user through a simple interface. This automated approach reduces manual effort, improves accuracy, and helps in faster and fairer selection of beneficiaries.

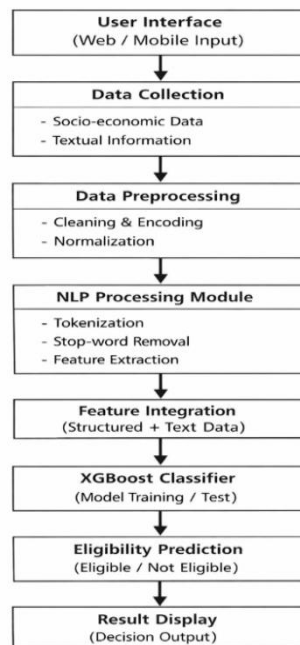


Fig.1. 1. System Architecture

B. Key Features

The proposed system includes the following key features:

- **ML-Based Prediction:** Uses the XGBoost algorithm to analyze socio-economic data and predict eligibility.
- **NLP Integration:** Processes text data like scheme details and applicant information.
- **Data Preprocessing:** Cleans and prepares data for better model performance.
- **Automated Decision:** Automatically determines eligibility, reducing manual work.
- **User-Friendly Design:** Simple interface suitable for rural users.

C. Workflow

The system works in a step-by-step process to predict eligibility for welfare schemes. First, it collects applicant details such as income, age, occupation, education, and family background, along with scheme descriptions and eligibility



rules. Next, the data is cleaned by handling missing values, correcting errors, and converting categories into a suitable format. NLP techniques are used to extract useful information from text data.

After preprocessing, the dataset is divided into training and testing sets. The XGBoost model is trained using the training data to learn patterns and then tested to check its accuracy. Once the model is ready, it can predict whether a new applicant is eligible based on their details.

The result is displayed instantly through a simple interface. This system reduces manual work, speeds up the process, and helps ensure that benefits reach the right people more accurately.

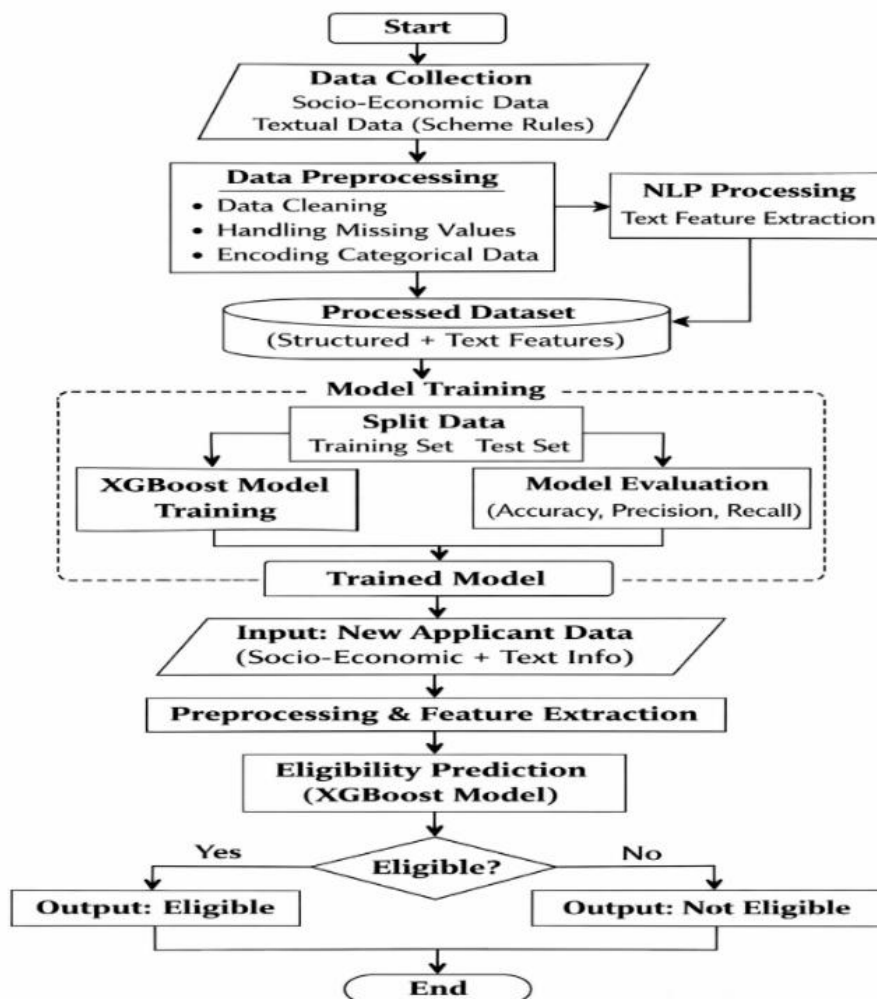


Fig.2. Workflow Architecture

V. IMPLEMENTATION DETAILS

In this project, a machine learning-based approach is used to predict eligibility for government welfare schemes. First, the required data is collected, including details such as age, income, occupation, education, family size, and location of applicants. Along with this, information about different welfare schemes, like eligibility rules and income limits, is also gathered from available datasets. After collecting the data, it is cleaned by removing duplicate entries and handling missing values. Categorical data is converted into numerical form so that it can be used by the model. For handling text data, such as schemes of common words, TF-IDF is applied.



Next, the important features are selected to improve the performance of the system. The dataset is then divided into training and testing parts. The XGBoost algorithm is used to train the model so that it can learn patterns from the data and predict whether an applicant is eligible or not. The model performance is checked using measures like accuracy, precision, recall, and F1-score. Finally, the trained model is connected to a simple web application built using HTML, CSS, JavaScript, and Python Flask. Users can enter their details, and the system quickly shows the result. This method reduces manual work, improves accuracy, and makes the process faster and more reliable.

A. Technologies Used

The system is developed using Python along with machine learning, NLP, and web technologies. Libraries like NumPy and Pandas are used for data processing, while NLTK and Scikit-learn handle text analysis using NLP techniques. The XGBoost algorithm is used for predicting eligibility due to its high accuracy. Flask is used for backend development, and HTML, CSS, and JavaScript are used to create the user interface. Data is stored in CSV files or MySQL, and tools like Jupyter Notebook and VS Code are used for development and testing.

VI. OUTPUT SCREENSHOTS

The system works in a simple and user-friendly way. First, the user enters details such as name, family information, occupation, and education through the website. The system collects this data and creates a complete profile. If the user provides any text input, NLP techniques are used to understand and process the information. After processing the data, the XGBoost model analyzes the applicant's details and predicts whether the person is eligible for a government welfare scheme or not. The result is then displayed clearly on the screen as "Eligible" or "Not Eligible." In some cases, the system may also suggest other suitable schemes. Overall, the process is fast, simple, and easy for users.

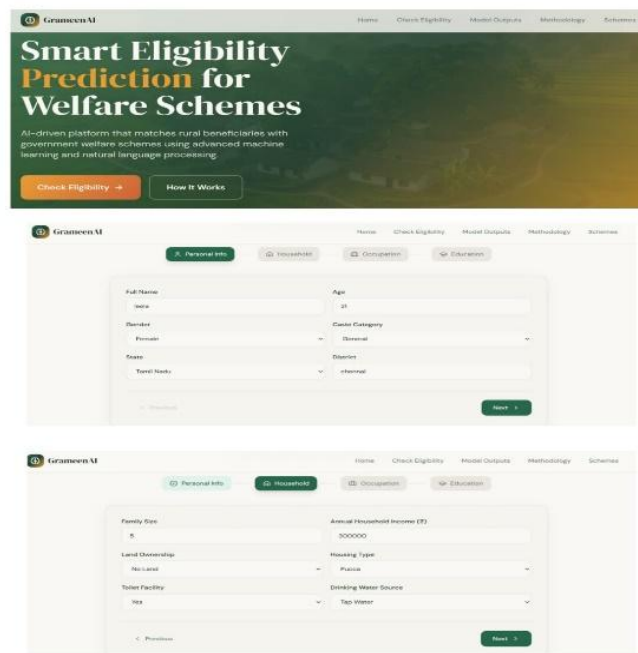


Fig. 3 .1 User Input Interface and Model Output Screens of Welfare Eligibility Prediction System

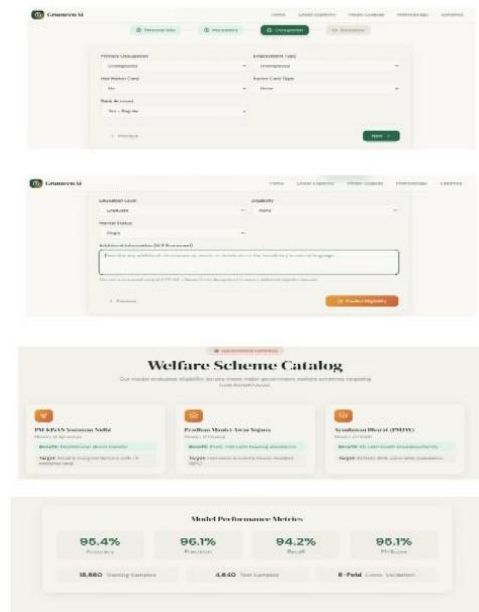


Fig. 3.2 System Output and Performance Visualization

A. Performance Metrics

Performance metrics are used to evaluate how well the proposed system predicts beneficiary eligibility. In this project, the performance of the XGBoost classification model is measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics help determine how effectively the model identifies eligible and ineligible applicants.

Accuracy measures the overall correctness of the model. It represents the percentage of total predictions that are classified correctly by the system.

Precision indicates how many of the applicants predicted as eligible are actually eligible. A higher precision value means the model makes fewer incorrect positive predictions.

Recall measures the ability of the model to correctly identify all eligible applicants from the dataset. It shows how well the system detects true beneficiaries.

The F1 score is the harmonic mean of precision and recall. It provides a balanced evaluation of the model, especially when the dataset contains uneven class distribution.

These metrics help analyze the reliability and efficiency of the proposed eligibility prediction system. A higher value in these metrics indicates better performance of the machine learning model in predicting the correct beneficiary status.

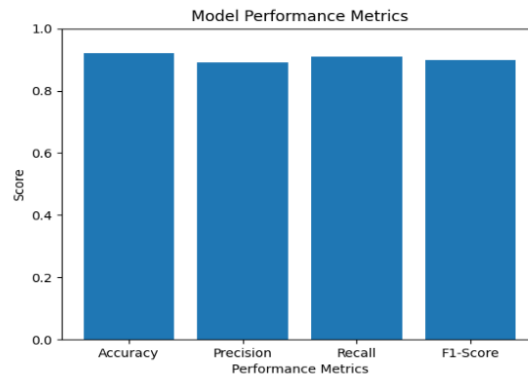


Fig. 5 Graphical Representation of Performance Metrics

D. Challenges

Building the beneficiary eligibility prediction system wasn't easy. Data availability and quality posed some serious headaches. We struggled to gather complete, reliable socio-economic data for training—many datasets had holes, inconsistencies, and messy entries. Cleaning that up took time and careful attention. Working with unstructured text was another sticking point. Since we relied on NLP to process scheme descriptions and applicant submissions, we had to wrangle these scattered bits of text into something the machine learning model could actually use. That meant a lot of extra preprocessing and feature extraction. Feature selection mattered, too. With dozens of attributes like income, occupation, education, and family size, we had to figure out which ones really influenced eligibility. Getting that right helped the model perform better. Training and tuning the model brought its own set of challenges. The XGBoost algorithm offers a sprawling set of parameters—finding the right mix wasn't straightforward. We had to experiment and tweak settings to prevent overfitting and boost accuracy. Then there was the interface. Designing something simple and user-friendly was crucial, especially for rural users who aren't always comfortable with technology. We needed a system that made sense to them but still gathered all the necessary information. Despite everything, we managed to pull it off. Careful data preprocessing, thoughtful feature engineering, and smart model optimization kept the system running smoothly and delivered accurate predictions for eligibility.

E. Discussion

We built this beneficiary eligibility prediction system to make it easier—and fairer—to figure out who qualifies for government welfare programs. Instead of relying only on time-consuming manual checks, the system taps into a variety of socio-economic data: personal details, family info, job, education, and more. It doesn't just stop at structured info—using Natural Language Processing (NLP), it handles free-text descriptions and scheme guidelines. XGBoost, a powerful machine learning algorithm, brings it all together to predict eligibility. When we tested the system, it had solid accuracy. XGBoost isn't just a buzzword—it really captures the complex connections hidden in the data, which usually trip up traditional methods. You see the improvement not just in numbers, but in the actual quality of predictions, leaving old manual shortlists behind. NLP pulls its weight too, sifting through unstructured information that otherwise ends up overlooked. By extracting richer features, the model gets a clearer picture of each applicant, leading to sharper, more relevant decisions. One big advantage: automation drastically cuts down on human bias and mistakes. Manual checks always leave room for error or inconsistency, and people don't always agree. In contrast, the data-driven approach of this system brings transparency and consistency—every applicant gets evaluated by the same rules. So, the takeaway is clear. Machine learning isn't just a theoretical upgrade here; it's a practical tool for making beneficiary selection faster, fairer, and more robust. With good data and routine updates, this approach keeps getting stronger—and as a result, government welfare programs can serve the right people, more efficiently and more justly, in the real world.

VII. CONCLUSION AND FUTURE WORK

A. Conclusion

This project introduces an AI-powered system that predicts who qualifies for government welfare programs. It draws on socio-economic data—personal details, family background, occupation, and education—to analyze each applicant's profile and determine eligibility. By combining Natural Language Processing (NLP) and the XGBoost algorithm, the system handles both structured and text data, strengthening its accuracy. Automating eligibility checks speeds up the



process, removes the bottleneck of manual verification, and makes outcomes more consistent and fair. XGBoost handles complex relationships among the various attributes, so the predictions aren't just fast—they're dependable. Users don't need technical expertise. The interface stays simple: they can enter their information and quickly find out if they qualify. This is particularly helpful in rural areas, where information about welfare schemes isn't always easy to access. In the big picture, this system shows how machine learning can make government welfare distribution smarter and more transparent. As the system evolves with more data and refinements, it stands to become a valuable tool for boosting efficiency and fairness in public welfare.

B. Future Work

The current system does a solid job predicting beneficiary eligibility, but there's plenty of room to grow. Expanding the dataset—making it bigger and more diverse—would push the model further. If we train it on data from a wider range of regions and scenarios, the predictions will hit closer to the mark and hold up under different conditions. There's also potential in layering in more machine learning models. Right now, the focus is mostly on XGBoost, but branching out—testing options like Random Forest or Neural Networks—offers a chance to see real performance gains. Comparing these models side by side helps pinpoint what works best. Making the system more accessible matters, too. A mobile app would be a game-changer, especially for people in rural areas who rely on smartphones. Pair that with a multilingual interface, and suddenly the service opens up to users in their own languages—no more struggling with language barriers. With ongoing work—better data, smarter models, and thoughtful system upgrades—this solution can become a key tool for government welfare program management and identifying the people who need support most.

REFERENCES

- 1) R. Kumar and S. Patel, "Machine Learning-Based Welfare Scheme Recommendation System," *Int. J. Innovative Research in Technology (IJIRT)*, vol. 12, no. 10, 2026.
- 2) Z. Khotanlou, K. Larson, and A.-H. Karimi, "Multi-Agent Algorithmic Recourse for Social Welfare Optimization," arXiv preprint, 2025.
- 3) P. M. Safwan et al., "Tech-Farm: Digital Empowerment and Welfare Scheme Integration Using AI," *Int. J. Scientific Research and Engineering Trends*, 2025.
- 4) S. Yfantidou et al., "Fairness in Machine Learning Systems: A Review," *IEEE Pervasive Computing / ACM IMWUT*, 2023.
- 5) 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
- 6) 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
- 7) 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
- 8) 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
- 9) 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.epr.2025.112428
- 10) 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
- 11) 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/aei-2013-0025.
- 12) 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.



- 13) 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.
- 14) 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
- 15) 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>
- 16) 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
- 17) IEEE Signal Processing Society, "AI in Signal and Data Science: Emerging Trends," IEEE JSTSP Special Issue, 2025.
- 18) K. Sivamayivelan et al., "AI-Based Chatbot System Using Large Language Models," Journal of Big Data, 2025.
- 19) Raja et al., "Generative AI-Based Decision Support System for Industrial Applications," Proc. IEEE IECON, 2025.
- 20) M. Iman et al., "A Review of Deep Transfer Learning and Recent Advancements," Technologies Journal, 2023.
- 21) IEEE SMC Society, "Machine Learning and NLP Applications in Social Systems," IEEE Trans. Computational Social Systems, 2025.
- 22) J. G. A. Barbedo et al., "Explainability and Privacy in AI-Based Systems," Computers and Electronics in Agriculture, 2026.
- 23) IEEE Region 4, "Machine Learning Applications in Public Decision Systems," IEEE Newsletter, 2024.
- 24) R. Das Gupta, "AI-Driven Decision Systems and Applications," Proc. IEEE Int. Conf., 2025.
- 25) IEEE Conference on Science and Technology, "Recent Advances in AI and Machine Learning Applications," 2025.
- 26) IEEE IES, "Generative AI Systems for Intelligent Decision Making," Proc. IEEE IECON, 2025.
- 27) SCAA Research Group, "Recent Advances in AI and Machine Learning Applications," 2026.