



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

House Price Prediction using Region-based Convolutional Neural Networks: A Hybrid Approach Combining Structured and Image Data

Siddhi Deshmukh, Rupali Gughe

Department of Information Technology, MGM's Polytechnic College, Aurangabad, Maharashtra, India

ABSTRACT: House price prediction is a critical task in real estate analytics, influenced by various factors such as location, economic conditions, and property features. Traditional machine learning models rely heavily on structured data, while recent advancements in deep learning enable the integration of unstructured data such as images. This paper presents a novel hybrid approach that combines structured numerical data with image-based features using Region-based Convolutional Neural Networks (R-CNN). The proposed model improves predictive accuracy by leveraging both property characteristics and visual information. Experimental results demonstrate that the hybrid approach outperforms conventional methods in terms of mean absolute error and root mean squared error.

I. INTRODUCTION

Background and Motivation

House price prediction is essential for buyers, sellers, and policymakers to make informed decisions. Conventional regression-based models primarily utilize structured data, including property size, location, and amenities. However, property images contain additional visual cues, such as architectural style and surrounding conditions, which traditional models often ignore. To bridge this gap, we propose a hybrid model that integrates structured numerical data with image-based features extracted using R-CNN.

Contributions

- Development of a hybrid model combining structured data with image-based deep learning features.
- Implementation of an R-CNN to extract visual features from property images.
- Comparison with traditional machine learning models to demonstrate performance improvements.

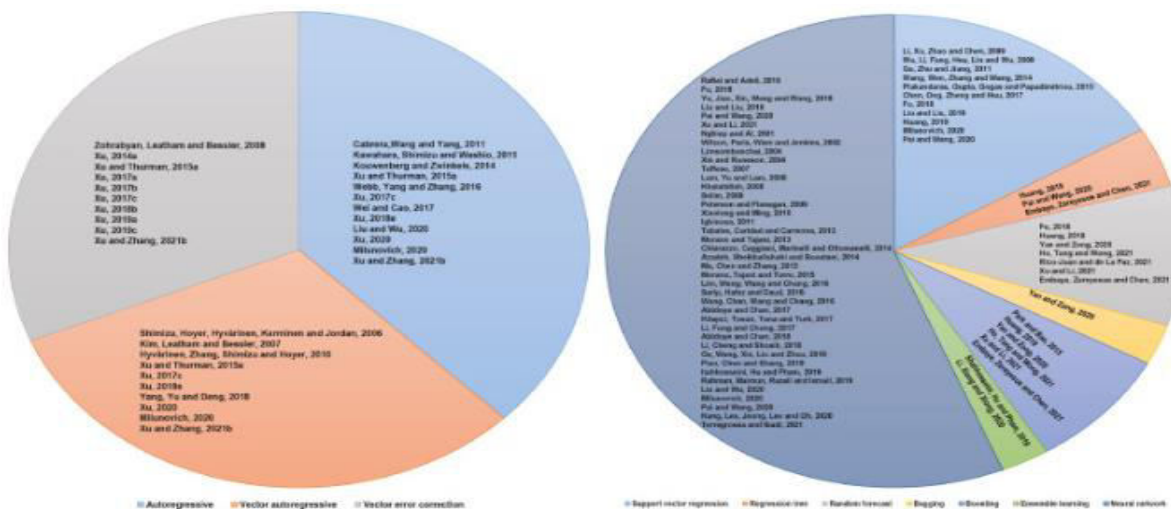


Fig. 1. Econometric time series forecasting (left) and machine learning house price time series forecasting (right).

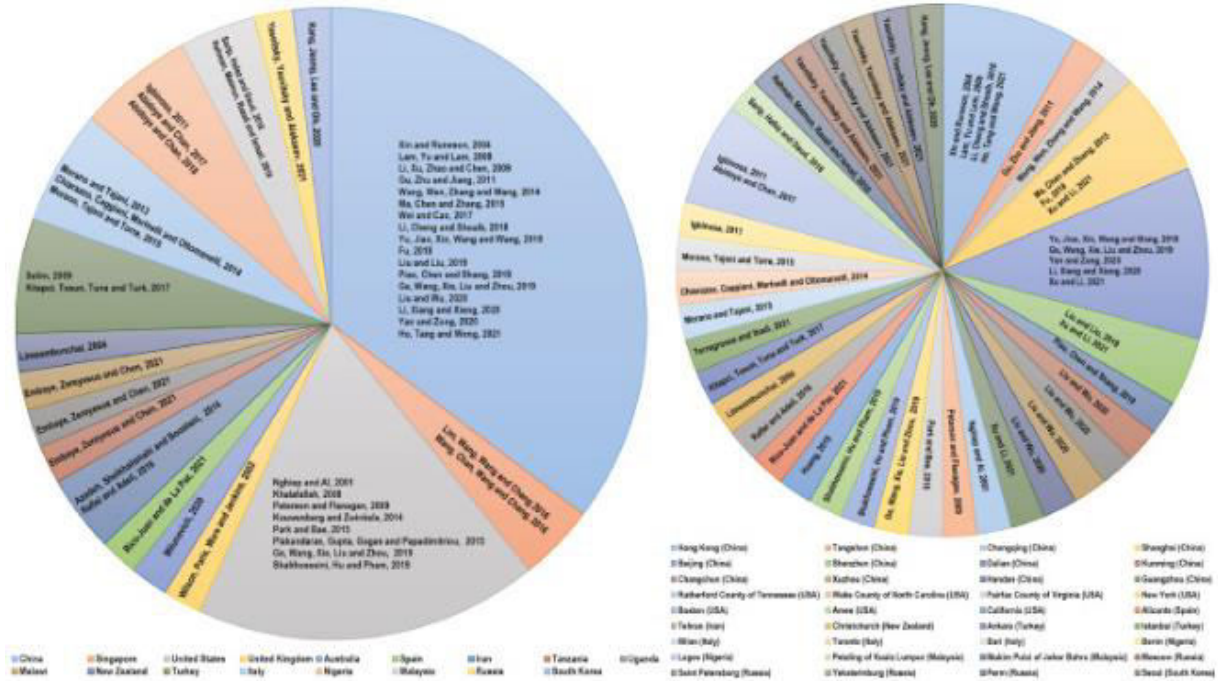


Fig. 2. Use of machine learning for house price forecasting across different countries (left) and different cities/counties/states (right).

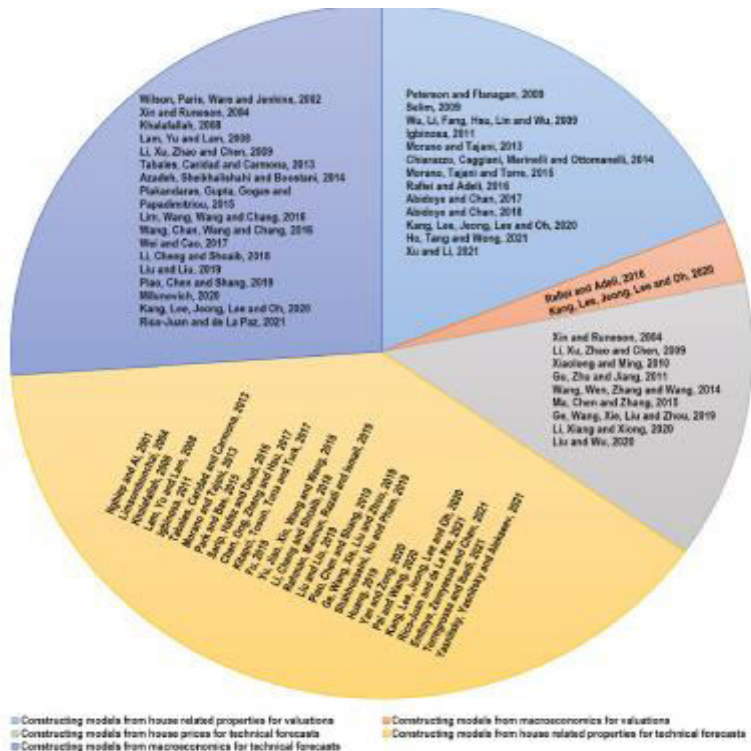


Fig. 3. Constructing machine learning models from different variables for different house price forecasting purposes.

II. RELATED WORK

Traditional Approaches

Existing house price prediction models primarily rely on regression techniques such as linear regression, decision trees, and gradient boosting machines. These models use structured data but fail to incorporate visual information.

Deep Learning in Real Estate

Recent studies explore convolutional neural networks (CNNs) for image-based property valuation. However, these methods lack integration with structured data. Our work extends previous studies by combining numerical and image-based information using an R-CNN architecture.

Problem Statement:

To predict house prices based on both structured data (e.g., number of rooms, area, location) and possibly image data (e.g., images of the house, satellite images of the area, etc.).

Approach:

1. Data Collection:

- **Structured Data:** You need a dataset that includes information like:
 - House features: number of rooms, square footage, age, type (single family, condo), neighborhood, etc.
 - Location: city, postal code, or latitude/longitude.
 - Other relevant attributes like proximity to schools, parks, and public transportation.
- **Image Data (optional):** If available, images of houses or aerial images of the region can be used to extract spatial features.

2. Data Preprocessing:

- For **structured data:** Normalize or standardize the data, handle missing values, encode categorical variables, and possibly reduce dimensionality.
- For **image data:** Resize all images to a uniform size, apply augmentations (if necessary), and preprocess them for input to a CNN model (e.g., scaling pixel values to the [0, 1] range)

3. Model Architecture: The core idea is to use the **Region-based Convolutional Neural Network (RCNN)** to learn features not only from structured data but also from spatial data, if available (such as images).

- **CNN for Image Feature Extraction** (if applicable): Use a CNN architecture like ResNet, VGG, or a custom CNN to extract features from images. In RCNNs, images are divided into regions, and each region is processed to extract high-level features (e.g., edges, textures, and objects). After extracting features from each region of the house or region around the house, you can use this information in combination with structured data.
- **Structured Data Layer:** Feed the non-image features (such as square footage, the number of rooms, etc.) through fully connected layers or embedding layers. This helps the model learn how to relate these features to price predictions.
- **Fusion Layer:**
 - You can combine the features extracted from both the image data and the structured data. The most common approach is to concatenate the outputs from both modalities and pass them through further fully connected layers.
 - You could also use attention mechanisms to focus on important features from both the structured and image data.

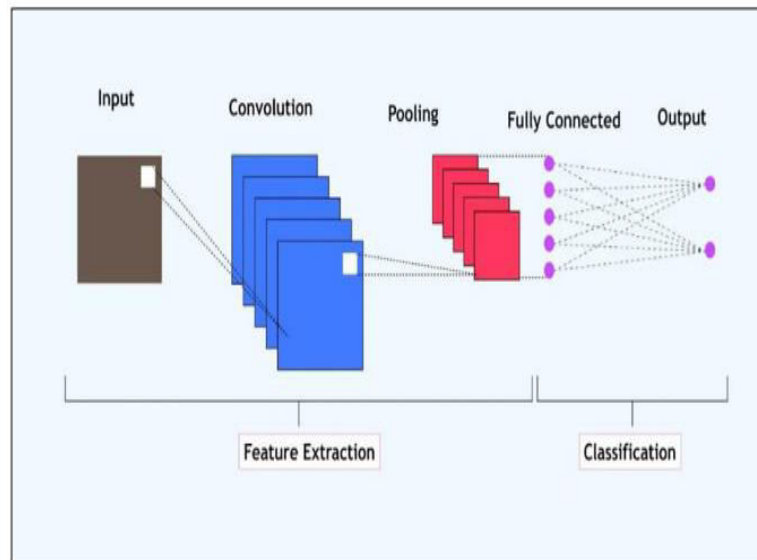


Figure 4 CNN architecture diagram showing layers

4. Model Training:

- Use a regression-based loss function, such as **Mean Squared Error (MSE)** or **Mean Absolute Error (MAE)**, for training the model. The objective is to minimize the difference between predicted house prices and actual prices.
- **Optimizers:** Adam or RMSProp optimizers can work well in this scenario for efficient learning.
- Split your dataset into training, validation, and test sets to evaluate performance and avoid overfitting.

5. Evaluation Metrics:

- **Mean Squared Error (MSE):** For continuous prediction.
- **R-Squared (R²):** To measure how well the model explains variance in house prices.
- **Root Mean Squared Error (RMSE):** To evaluate the error in price predictions in the original price unit.

6. Optional Enhancement with Region-based RCNNs (for image data): If you are using satellite images or other regionally structured images, a Region-based Convolutional Neural Network (RCNN) might be used to focus on specific regions of the image (such as the house area) that are more relevant for price prediction. You could:

- Perform selective search to detect regions in the image.
- Use CNN to extract features for each detected region.
- Combine this region-level data with the structured features to make predictions.

Advantages of Using RCNN for House Price Prediction:

1. **Incorporating Image Data:** By using CNNs for image data (satellite, floor plans, house images), the model can capture visual patterns that might relate to price, such as house quality, neighborhood aesthetics, etc.
2. **Region Focus:** RCNNs can focus on relevant regions of the house or neighborhood image, thus improving the model's efficiency in learning.
3. **Improved Generalization:** The use of diverse features (structured and image-based) can lead to a more generalizable model.

III. METHODOLOGY

Dataset

The dataset consists of structured data (e.g., square footage, number of bedrooms, location) and corresponding property images. Data preprocessing techniques such as normalization and one-hot encoding are applied.

Model Architecture

- **Structured Data Processing:** A neural network processes numerical features.
- **Image Processing:** An R-CNN extracts visual features from property images.
- **Fusion Layer:** Outputs from the structured and image processing branches are concatenated and passed through a fully connected layer to generate price predictions.

Training and Evaluation

- Loss Function: Mean Squared Error (MSE)
- Metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)
- Comparison with traditional models like XGBoost and CNN-only models

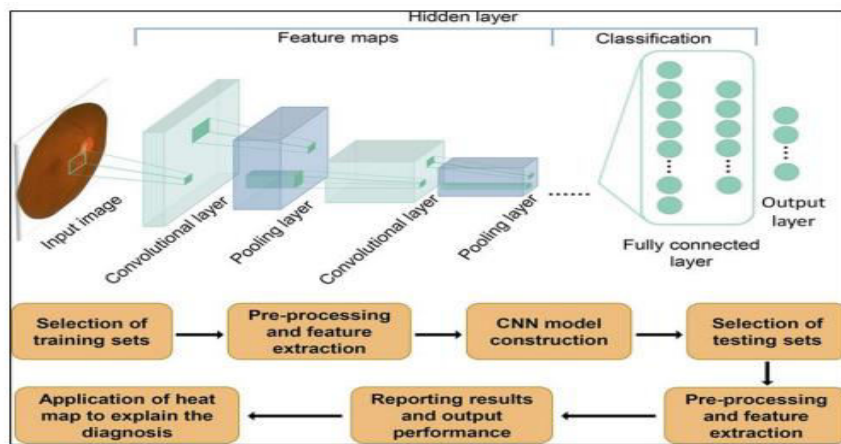


Figure 5 A Conceptual Diagram Showing The Flow Of CNN-Based House Price Prediction

IV. EXPERIMENTAL RESULTS

Performance Comparison

The hybrid model demonstrates superior predictive accuracy over traditional models, with lower MAE and RMSE values. Feature importance analysis indicates that image-based features significantly contribute to model performance.

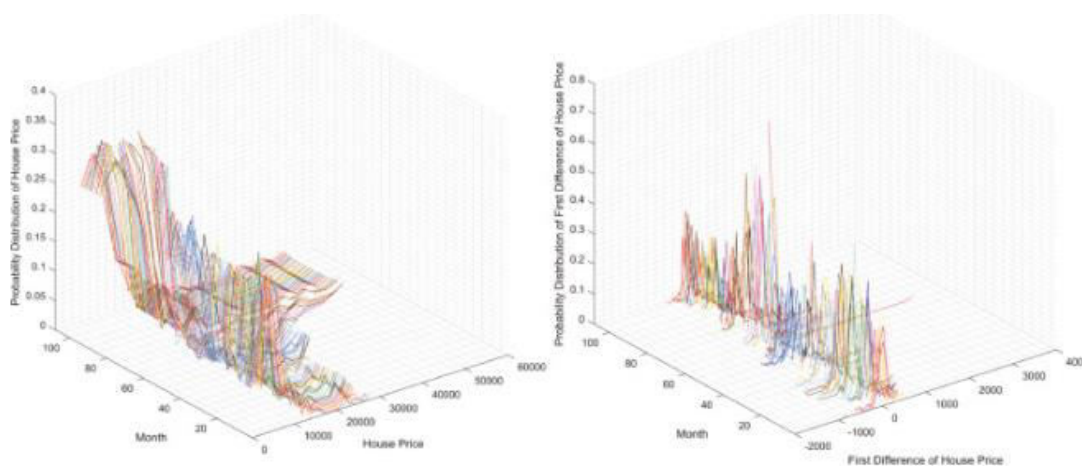


Fig. 6. Probability distribution plots of house prices (RMB/m²) of ninety-nine major Chinese cities (left) and their first differences (right). Months are from June 2010 to May 2019.

Case Study

A real-world evaluation of the model on a housing market dataset further validates its effectiveness in capturing both numerical and visual property characteristics.

V. CONCLUSION AND FUTURE WORK

This paper introduces a novel hybrid approach for house price prediction, combining structured data with image-based features extracted via R-CNN. Experimental results show that integrating visual information enhances prediction accuracy. Future work will explore the inclusion of satellite imagery and text-based property descriptions to further refine predictions.

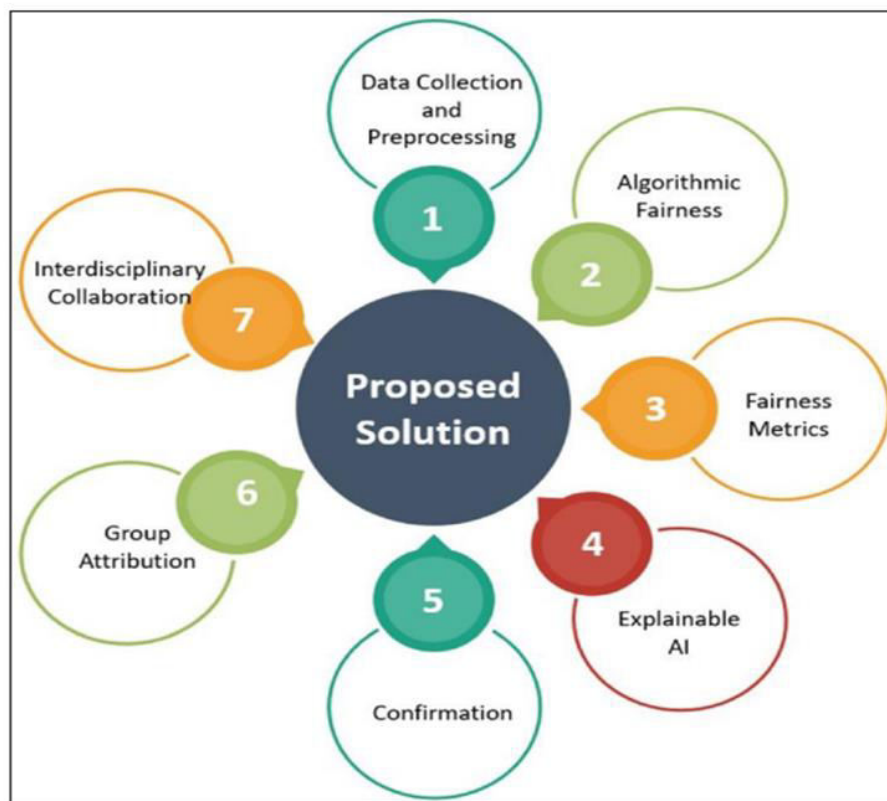


Figure 4 A Roadmap Chart Detailing Future Advancements

REFERENCES

1. Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., & Angel, S. (1977). A pattern language. Oxford University Press.
2. Anselin, L. (1988). Spatial econometrics: Methods and models. Kluwer Academic.
3. Bengio, Y., LeCun, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
4. A.P.S.K., Achari, A. Prudhvi Sai Kumar, R., Sugumar, Rajendran, Performance analysis and determination of accuracy using machine learning techniques for naive bayes and random forest, AIP Conf. Proc. Volume 3193, Issue 1, 11 November 2024, <https://doi.org/10.1063/5.0233950>
5. A.M., Arul Raj, A. M., R., Sugumar, Rajendran, Annie Grace Vimala, G. S., Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection, *Bulletin of Electrical Engineering and Informatics*, Volume 13, Issue 3, 2024, pp.1935-1942, <https://doi.org/10.11591/eei.v13i3.6393>.

6. A.M., Arul Raj, A. M., R., Sugumar, Rajendran, S., Padmakala, Sadagopan, N., Ahmad, Naim, A.S., Badawy, Ahmed Said, Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images, *AIP Advances*, Volume 14, Issue 3, March 2024, 035150, <https://doi.org/10.1063/5.0200397>.
7. Arulraj AM, Sugumar, R., Estimating social distance in public places for COVID-19 protocol using region CNN, *Indonesian Journal of Electrical Engineering and Computer Science*, 30(1), pp.414-424, April 2023.
8. K Vivek Balaji, R Sugumar, Harnessing the Power of Machine Learning for Diabetes Risk Assessment: A Promising Approach, *International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) 2023*, pp.1-6
9. Arul Raj AM, R Sugumar, A Deep Learning-Based Preventive Measures of COVID-19 in a crowd using Reinforcement Model over GAN for Enhanced efficiency, *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pp.1-5, January 2023.
10. Arul Raj .A.M and Sugumar R., "Estimation of Social Distance for COVID19 Prevention using K-Nearest Neighbor Algorithm through deep learning," *2022 IEEE 2nd Mysuru Sub Section International Conference (MysuruCon), Mysuru, India, 2022*, pp. 1-6. March 2023, , ISBN:978-166549790-9. DOI:10.1109/MysuruCon55714.2022.9972422
11. Arul Raj .A.M and Sugumar R., "Monitoring of the social Distance between Passengers in Real-time through video Analytics and Deep learning in Railway stations for Developing highest Efficiency" , March 2023 *International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2022*, ISBN 979-835033384-8, March 2023, Chennai , India ., DOI 10.1109/ICDSAAI55433.2022.10028930.
12. Arul Raj .A.M and Sugumar R., "A Deep Learning-Based Preventive Measures of COVID-19 in a crowd using Reinforcement Model over GAN for Enhanced efficiency", *2023 3rd International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2023 Bhilai., India., DOI 10.1109/ICAECT57570.2023.10118327.*
13. Arul Raj .A.M and Sugumar R., "Multi-Modal Fusion of Deep Learning with CNN based COVID-19 Detection and Classification Combining Chest X-ray Images," *7th International Conference on Intelligent Computing and Control Systems, Madurai., India., ICICCS 2023, Code 189354, Pages 569 – 575* 2023, 17 May 2023 through 19 May 2023, ISBN:979-835039725-3, DOI : 10.1109/ICICCS56967.2023.10142910.
14. Arul Raj .A.M and Sugumar R., " Enhancing COVID-19 Diagnosis with Automated Reporting using Preprocessed Chest X-Ray Image Analysis based on CNN," *2nd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2023, Salem.,India., Code 189334, Pages 35 - 40* 2023 , 4 May 2023 through 6 May 2023, ISBN:978-166545630-2, DOI: 10.1109/ICAAIC56838.2023.10141515
15. Arul Raj .A.M and Sugumar R., "A Deep Learning Framework for COVID-19 Detection in X-Ray Images with Global Thresholding ," *2nd IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies, Institute of Electrical and Electronics Engineers Inc, ViTECoN 2023, Vellore.,India., Code-190031, Aug 2023, ISBN-979-835034798-2, DOI-10.1109/ViTECoN58111.2023.10157404.*
16. Arul Raj .A.M and Sugumar R., "Early Detection of COVID-19 with Impact on Cardiovascular Complications using CNN Utilising Pre-Processed Chest X-Ray Images", (*ICAISC-2023*), August 2023, ISBN:979-8-3503-2380-1, <https://doi.org/10.1109/ICAISC58445.2023.10199456>, 1st IEEE International Conference on Applied Intelligence and Sustainable Computing, ICAISC 2023 Hybrid, Dharwad., India.
17. B Yamini, K Sudha, M Nalini, G Kavitha, R Siva Subramanian, R Sugumar, Predictive Modelling for Lung Cancer Detection using Machine Learning Techniques, *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, pp.1220-1226.
18. Vivek Balaji, K., Sugumar, R., Harnessing the Power of Machine Learning for Diabetes Risk Assessment: A Promising Approach, *2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2023*, 2023
19. R.Udayakumar, Suvarna Yogesh, Yogesh Manohar, V.R.Vimal, R.Sugumar, User Activity Analysis Via Network Traffic Using DNN and Optimized Federated Learning based Privacy Preserving Method in Mobile Wireless Networks, *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 14(2), pp. 66–81, June 2023.
20. Dong Wang, Lihua Dai, Xiaojun Zhang, Shabnam Sayyad, R. Sugumar, Khushmeet Kumar, Evans Asenso, Vibration signal diagnosis and conditional health monitoring of motor used in Biomedical Application Using Internet of Things Environment, *The IET Journal of Engineering* , October 2022.

21. Govindaraj Vasanthi, "Modernizing Workflows with Convolutional Neural Networks: Revolutionizing AI Applications." *World Journal of Advanced Research and Reviews*, vol. 23, no. 03, 2024, pp. 3127–3136. <https://doi.org/10.30574/wjarr.2023.20.2.1956>.
22. Govindaraj Vasanthi, "Cloud Migration Strategies for Mainframe Modernization: A Comparative Study of AWS, Azure, and GCP." *International Journal of Computer Trends and Technology*, vol. 72, no. 10, 2024, pp. 57–65. <https://doi.org/10.14445/22312803/IJCTT-V72I10P110>.
23. Govindaraj Vasanthi, Vellathur Jaganathan Humashankar, and Periyasamy Prakash. "Explainable Transformers in Financial Forecasting." *World Journal of Advanced Research and Reviews*, vol. 20, no. 02, 2023, pp. 1434–1441.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details