iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82696

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Dog Behaviour Analysis

Ajay Kumar Meena
Centre of AI, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India

Abstract: Dogs show a high degree of heterogeneity in their behavior patterns because of the close association between humans and dogs in many capacities – domesticated pets, assistance animals and experimental subjects. Heterogeneity of dogs' behavior can be attributed both to intrinsic and extrinsic reasons including, among others, breed, genetics, age, environment, human interaction, training programs. Accurate dog behavior prediction is becoming increasingly important from the point of view of animal welfare, safety, and training efficiency. In this study, we attempt to examine different machine learning models that can help predicting dog behavior patterns based on structured datasets. By analyzing certain behavioral and contextual factors, this study intends to evaluate the efficacy of models such as Artificial Neural Networks and Random Forests. The general idea is to find some patterns and predictors, which can work with any individual dogs and allow predicting their behavior. Behavioral tests and observation were popular tools for evaluating temperaments in animals, however, the results of these practices are often subjective, inconsistent and lack scalability. Machine learning offers an innovative approach to behavior pattern analysis which involves large volumes of data processing. Supervised machine learning classifiers will be trained and tested by using certain evaluation metrics including accuracy, precision, recall, F1-score. From the results, it can be noted that there is no model that is superior to all other models. Every model has its strengths based on the behavior class it will be applied to and the type of features used. This highlights the significance of choosing the best models and developing suitable features in predicting behaviors. In conclusion, this study adds value to computational ethology by proposing the use of AI in behavioral science.
Keywords: Sensor-Based Behavior Prediction, Time-Series Classification, Dog Behavior Analysis, Long Short-Term Memory (LSTM), Machine Learning.

I. INTRODUCTION

The dogs (Canis lupus familiaris) are considered to be domesticated and therefore wild animals that live alongside humans. Through time, they have come to be extremely close to humans, and their continual evolution alongside humans is one of the primary reasons why they exist in human society. They have also been used in far different ways such as being great friends or perfect companions in human households. Besides, the roles that dogs occupy in society depend largely on their nature. Due to their intelligence and responsive characters, the dogs are involved in different activities like animal behavioral control, having social sensitivity, and exercising cognitive flexibility through their very consistent and regular performance. Dogs also display a lot of variety in their behavior. Breeds, genetics, age, sex, training history, environment, and early socialization experiences are some of the factors that have a major impact on how individual dogs act in different situations. This diversity in behavior offers both advantages and drawbacks. For instance, some dogs will be under the high-stress or social environmental category and will perform very well while others will be aggressive or anxious or may fear the situation and hence nothing will be able to convince them to work or they might cause a problem. Evaluating and anticipating dog behavior is a matter of increasing significance in the fields of animal science, veterinary behavior, rescue shelters, and working dog organizations. Conventional behavioral tests, including temperament testing cognitive testing, and assessment by trainers, offer valuable insight. However, conventional approaches have their drawbacks.
Generally, they are highly subjective, labor-intensive, and significantly affected by inter-observer variation. Furthermore, the lack of standardization in the process of conducting behavior tests has contributed to inconsistencies in behavior measurement and interpretation.

II. TYPE STYLE AND FONTS

1) Font Style: Times New Roman
2) Title Font Size: 24 pt, Bold, Uppercase, Center Aligned
3) Author Details: 10 pt, Center Aligned
4) Main Headings: 14 pt, Bold
5) Subheadings: 10 pt, Bold
6) Body Text: 10 pt, Justified Alignment
7) Line Spacing: Single

8) Page Size: A4
9) Margins:
   - Left: 0.51”
   - Right: 0.51”
   - Top: 0.75”
   - Bottom: 0.75”
10) Figure Caption: 10 pt, below figure, centered
11) Table Caption: 9 pt, above table, centered

III. TERMINOLOGIES

Table I
Abbreviations and their Full Forms

Abbreviation	Full Form
ANN	Artificial Neural Network
LSTM	Long Short-Term Memory
RF	Random Forest
AI	Artificial Intelligence
ML	Machine Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
CSV	Comma-Separated Values

IV. DATASET DESCRIPTION AND PREPROCESSING

The dataset utilized in this research has been obtained from "Animal Behavior Analysis," which is an open-source dataset available on Kaggle and curated by Arashnia (2022). The dataset used is intended explicitly for machine learning algorithms for the purpose of detecting dog behavior from motion sensor signals. The measurement of data is performed using a device called an accelerometer and a gyroscope that are placed on the dog's back and neck regions. A single data point represents one time-step measurement in milliseconds.

For both regions, back and neck, there are measurements taken using two sensors, namely, the accelerometer and gyroscope, on three axes. Additional data associated with each sample include:

1) Dog ID: Unique identification number of a particular dog.
2) Test Number: Unique identification number of an experiment.
3) Time (t_sec): Timestamp of the sample in seconds.
4) Task: Controlled action (walking task).
5) Behavior_1 and Behavior_2: Action labels (walking, eating).

The multimodal label scheme enables analysis of single or multiple behaviors simultaneously, making the classification problem realistic but complex.

Table II
Data Fields Used in the Dataset

Feature	Description
t_sec	Time elapsed in seconds.
ABack_x/y/z	Accelerometer readings from the dog’s back on x, y, z axes.
ANeck_x/y/z	Accelerometer readings from the dog’s neck.
GBack_x/y/z	Gyroscope data from the dog’s back.
GNeck_x/y/z	Gyroscope data from the dog’s neck.

Task	The designed task during which data was collected.
Behavior_1	Primary behavior label for supervised learning.
Behavior_2	Secondary behavior label (optional for multi-label classification).
TestNum	Trial number under which the data was collected.
DogID	Identifier for the dog being observed.

V. MACHINE LEARNING MODELS

A. Random Forest

Random Forest is an ensemble learning algorithm that is commonly used and outputs the majority vote for classification tasks after constructing a collection of decision trees during the training phase. The main thought behind Random Forest is that weak learners (individual decision trees) can potentially turn into strong classifiers through the right combination of techniques. A tree is made up of samples from the training set, and every split a random subset of features is considered (a process known as bootstrap sampling). This randomness adds diversity to the trees and allows the whole model not to be as much affected by overfitting as it would with a single decision tree.

Random Forest is primarily recognized for its strength, simplicity of implementation, and capacity to assess feature importance, which is a great tool for piecing together variables that have the highest say in the model's predictions.

Strengths:

- Overfitting is of no concern: The averaging process saves one from the risk of getting the model fit to the noise in the data.
- Very suitable for high-dimensional data: Processing a huge number of features requires no much preprocessing.
- Grants importance to features: Aids in recognizing the most influential inputs for the model's predictions.
- Good on structured/tabular data: Particularly on datasets such as sensor readings, the model is highly effective.

Limitations:

- Harder to interpret than a single decision tree: Though quite a bit more stable, the “forest” is not as easy to visualize.
- The predicting time can be longer on large models: Especially when a great number of trees are used.
- Doesn't recognize temporal dependencies: Each input gets treated as a totally new entity, which is one of the reasons it is not suitable for sequence modeling unless proper engineering is done.

B. Artificial Neural Network (ANN)

Artificial neural networks are biological neural networks mimicking the structure and function of the human brain. They are comprised of layers of interconnected nodes (neurons). ANN converts data inputs through weighted connections and activation function outputs. The input layer accepts the data input (sensor data), one or more layers of hidden layers transform the data, and the output layer delivers a classification label output.

One significant advantage of ANNs is their capability to discover intricate relationships in complex and nonlinear datasets. They learn by adjusting connection weights using a training algorithm based on the minimization of an objective loss function, backpropagation of errors throughout the network, and updating the network. As a result of their learning process, artificial neural networks are capable of discovering intricate relationships within high dimensional data, including time-series motion sensor data.

Although highly effective, artificial neural networks can be considered a black box model, meaning that decision-making processes are opaque. Therefore, careful tuning of parameters such as learning rate, architecture, and regularization becomes crucial.

Strengths:

- Can model complex and non-linear relations: Best for tasks where simpler models fail.
- Scalable architecture: Neural networks and layers can be modified depending on the complexity of the task.
- Effective in large-scale data sets: More data leads to better learning.
- Versatile applications: Effective for classification, regression, time-series, etc.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Limitations:

- Acts as a "black box": Difficult to understand how and why certain decisions are made.
- Needs much tuning: The model performance largely relies on parameters like learning rate, number of layers, etc.
- Sensitive to noise or imbalance in data: May result in overfitting unless the model is regularized properly.

C. Long Short-Term Memory (LSTM)

Long Short-Term Memory is a deep learning algorithm that is applied for processing sequential or timeseries information. This is a specific kind of RNN architecture, which has the ability to retain essential information about inputs from earlier stages for a longer period of time. This feature makes LSTM especially valuable in applications that require using information from previous data for predicting further actions and events, such as movement or behavior prediction.

The LSTM architecture consists of memory blocks and various gates controlling information flow in the network. Such mechanisms are responsible for defining which information needs to be retained or discarded, thus making the algorithm capable of finding complex patterns in sequence of data.

In this study, LSTM will be employed to examine the series of readings from motion sensors attached to dogs. This is because canine behavior is not an event; it is a continuous process that happens throughout the day. LSTM will help in detecting any variations in movements and identifying actions like walking, feeding, sleeping, and smelling using time-series data.

Even though LSTM networks offer superior accuracy in predicting sequences, they need more processing power than conventional machine learning algorithms. Besides, their effectiveness depends on parameter adjustment to achieve maximum efficiency.

Strengths:

- Works very well with time-series and sequential data.
- Can remember important information from previous time steps.
- Effective in recognizing complex behavior patterns.
- Helps overcome short-term memory limitations found in traditional RNNs.

Limitations:

- Requires more training time and computational resources.
- More complex to implement compared to basic machine learning models.
- Needs a larger amount of data for better performance.
- Sensitive to parameter tuning and model configuration.

VI. RESULTS & INTERPRETATION

```
=== Classification Report ===
                 precision    recall  f1-score   support

Carrying object       0.00      0.00      0.00        16
       Drinking       0.31      0.06      0.10        67
         Eating       0.00      0.00      0.00       192
    Lying chest       0.61      0.78      0.69      1227
         Pacing       0.00      0.00      0.00        85
        Panting       0.54      0.46      0.50       929
        Playing       0.61      0.63      0.62      1014
        Shaking       1.00      0.07      0.12        45
        Sitting       0.62      0.40      0.49       569
       Sniffing       0.81      0.94      0.87      1157
       Standing       0.52      0.43      0.47       484
       Trotting       0.51      0.50      0.50       850
        Walking       0.59      0.77      0.67       865

       accuracy                           0.62      7500
      macro avg       0.47      0.39      0.39      7500
   weighted avg       0.59      0.62      0.60      7500

Accuracy: 0.62
Precision (macro): 0.47056542092069126
Recall (macro): 0.38800854092276976
F1-Score (macro): 0.387093217565443
```

Fig. 1 Report of ANN

```
=== Random Forest Classification Report ===
                precision    recall  f1-score   support

Carrying object      0.00      0.00      0.00        16
       Drinking      0.90      0.39      0.54        67
         Eating      0.46      0.06      0.11       192
    Lying chest      0.79      0.90      0.84      1227
         Pacing      0.00      0.00      0.00        85
        Panting      0.70      0.71      0.70       929
        Playing      0.74      0.70      0.72      1014
        Shaking      1.00      0.49      0.66        45
        Sitting      0.75      0.73      0.74       569
       Sniffing      0.90      0.96      0.93      1157
       Standing      0.82      0.58      0.68       484
       Trotting      0.75      0.81      0.78       850
        Walking      0.71      0.89      0.79       865

       accuracy                          0.77      7500
      macro avg      0.66      0.56      0.58      7500
   weighted avg      0.76      0.77      0.76      7500

Accuracy: 0.7730666666666667
Precision (macro): 0.6557258560973444
Recall (macro): 0.5554358419230613
F1-Score (macro): 0.576532638247241
```

Fig. 2 Report of Random Forest

```
accuracy: 0.7395 - loss: 0.5231 - val_accuracy: 0.7498 - val_loss: 0.5117

accuracy: 0.7535 - loss: 0.5021 - val_accuracy: 0.7678 - val_loss: 0.4818

accuracy: 0.7813 - loss: 0.4538 - val_accuracy: 0.7892 - val_loss: 0.4372

accuracy: 0.7917 - loss: 0.4299 - val_accuracy: 0.7913 - val_loss: 0.4230

accuracy: 0.7974 - loss: 0.4206 - val_accuracy: 0.8016 - val_loss: 0.4137

accuracy: 0.7992 - loss: 0.4161 - val_accuracy: 0.8021 - val_loss: 0.4160

accuracy: 0.8018 - loss: 0.4108 - val_accuracy: 0.8051 - val_loss: 0.4101

accuracy: 0.8075 - loss: 0.4049 - val_accuracy: 0.8099 - val_loss: 0.3989

accuracy: 0.8148 - loss: 0.3919 - val_accuracy: 0.8187 - val_loss: 0.3824

accuracy: 0.8279 - loss: 0.3699 - val_accuracy: 0.8527 - val_loss: 0.3370


accuracy: 0.8527 - loss: 0.3370
```

Fig. 3 Report on LSTM

However, from the experiment, it was evident that the three models successfully classified the dog's behavior with the help of sensor data; however, the performance varied between the models. For instance, the Random Forest model delivered consistent and relatively high accuracy results with an accuracy rate of around 77%. On the other hand, the accuracy of the ANN model was only about 62%, and it did not capture any temporal trends in the data set. Out of the three models used, LSTM had the highest accuracy rate of 85% with low loss values of 0.33. The ability of the model to learn from sequential time-series data enabled it to identify patterns in the behavioral data better compared to other models.

VII. CONCLUSION

The objective of the current research paper is to examine the efficiency of machine learning and deep learning methods used for predicting the behavior of dogs by employing information obtained from motion sensors using wearable devices. Several machine learning algorithms, such as Random Forest, ANN, and LSTM, were implemented and evaluated to measure their efficiency in terms of their ability to predict certain behavioral patterns, namely, walking, feeding, smelling, and sleeping.

LSTM proved to be the most efficient model as far as its accuracy is concerned (the figure is about 85%), while the loss was minimal (the value of 0.33). Such a high level of accuracy can be explained by the model's ability to find connections in time between the observations. As behavior is an ongoing process, LSTM was able to recognize better the movement patterns between different activities.

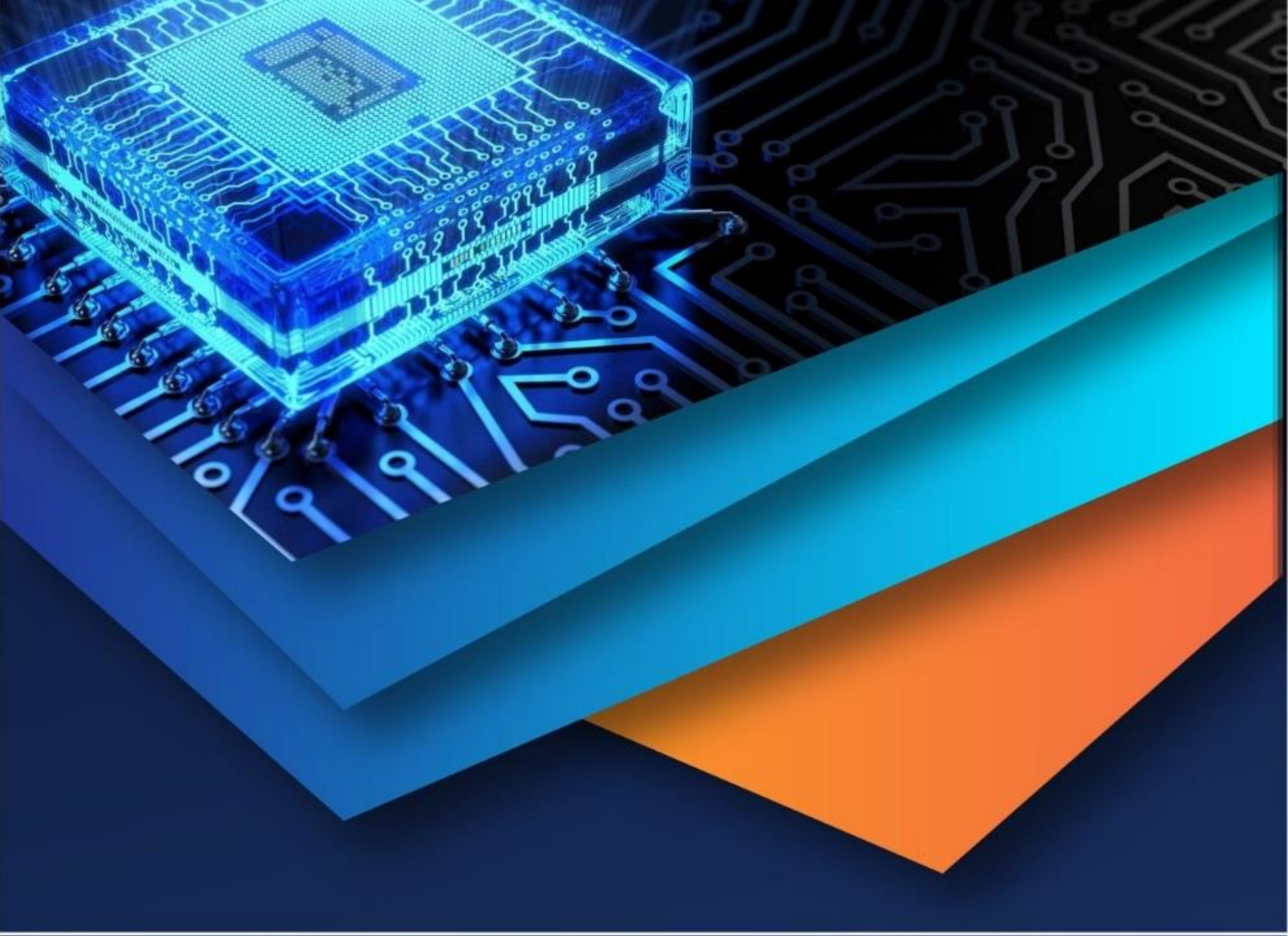
As for the Random Forest model, its performance was also impressive and resulted in stable and accurate predictions with high interpretability. This classifier efficiently processed structured sensor data and outperformed the ANN model in terms of accuracy. Nevertheless, the ANN model was able to find complex non-linear patterns in the dataset; however, its performance level was slightly worse than those of the Random Forest and LSTM models since it lacks temporal memory.

Thus, the results show that although the application of classical machine learning techniques can result in successful classification of dog behavior, deep-learning approaches prove to be more efficient for analyzing the time-series data. These findings prove that AI models can significantly contribute to the development of systems designed for monitoring and predicting canine behavior.

In conclusion, this paper demonstrates that the integration of sensor devices and AI models can facilitate behavioral analysis of dogs to a significant extent. Future research on the topic could include the use of bigger datasets, real-time analysis tools, and more efficient models like the GRU or CNN-LSTM ones.

REFERENCES

[1] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back- propagating errors,” *Nature*, vol. 323, pp. 533–536, 1986.
[3] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
[4] G. Ke et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 3146–3154.
[5] M. Florian and J. Kottferová, “Dog Behaviour Prediction Testing,” *International Journal of Veterinary Science*, vol. 11, no. 4, pp. 230–237, 2022.
[6] A. Arashnia, “Animal Behavior Analysis Dataset,” *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/arashnic/animal-behavior-analysis
[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
[8] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*, 3rd ed., Packt Publishing, 2019.
[9] F. Chollet et al., “Keras: Deep Learning for Humans,” GitHub Repository. [Online]. Available: https://keras.io
[10] scikit-learn documentation, “User Guide – Classification,” [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html
[11] TensorFlow Developers, “TensorFlow Documentation,” *Google Developers*. [Online]. Available: https://www.tensorflow.org
[12] R. J. Williams and D. Zipser, “A learning algorithm for continually running fully recurrent neural networks,” *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.
[13] J. Brownlee, “Gentle Introduction to Gradient Boosting,” *Machine Learning Mastery*, 2019. [Online]. Available: https://machinelearningmastery.com
[14] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556 [cs.CV], 2014.
[15] G. Hinton, “A Practical Guide to Training Restricted Boltzmann Machines,” *Technical Report*, University of Toronto, 2010.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)