iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76024

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 13 Issue XII Dec 2025- Available at www.ijraset.com

Comparative Study of Insider Threat Detection Models: Sentiment-Augmented Random Forest (SARF) Versus Isolation Forest, Autoencoder, and SVM

Neetha Natesh[1], Dr. Vidyarani H J[2], Darshan Gowda K J[3], Maheedhar C G[4], Tushar R Deshpande[5], Vijay Adithya J[6]
[1]*Assistant Professor, UG – Research Scholar, Programme of Information Science & Eng., School of Computer Science & Eng., Dr. Ambedkar Institute of Technology, Bengaluru, India*
[2]*Head Of Programme, Programme of Computer Science and BS, School of Computer Science & Eng., Dr. Ambedkar Institute of Technology, Bengaluru, India*

Abstract: ***Insider threats continue to be one of cybersecurity's most critical problems.Their presence is invisible amid normal user patterns, buried under what looks like ordinary user activity but carries hidden intent. Most anomaly detection systems stick to monitoring system behavior—logins, file transfers, network use—yet they miss something essential: the emotional undertone that often appears before a real incident unfolds. This study introduces the Sentiment-Augmented Random Forest (SARF), a compact and explainable model built to fuse behavioral data with emotional context drawn from user communications. SARF doesn't just watch what people do it listens to how they sound when they do it. By combining activity metrics with measures of sentiment polarity and subjectivity, it captures the subtle psychological drift that can signal risk before it turns into damage.***
We generated a synthetic dataset of 2000 simulated enterprise users to test the model, representing both normal and insider-like behavior. SARF reached an accuracy of 96.4% and an ROC-AUC score of 0.974. The results held up under statistical scrutiny, showing clear gains over baseline models. Feature importance plots make the model's decisions transparent, giving analysts a clear look into what drove each prediction. Beyond its technical results, SARF connects two worlds that rarely meet cybersecurity and human psychology. It's a step toward defense systems that don't just monitor data, but understand emotion, turning cybersecurity from a purely technical shield into something more perceptive, more human.
Index Terms: Insider threat detection, sentiment analysis, Random Forest, behavioral profiling, explainable AI, cybersecurity.

I. INTRODUCTION

Insider threats differ from external attacks in a very real way—the danger comes from people already inside the system, people who are trusted. They have legal access, so their actions rarely get recognized at first. Yet, long before an incident happens,
these individuals often show small shifts in how they behave or communicate—patterns that can be easy to overlook. Traditional anomaly detectors like Isolation Forest or Autoencoders do catch unusual technical activity, but they miss the linguistic and emotional signs hiding in everyday communication.
The idea behind SARF came from observing that human behavior itself leaves traces. Emotions such as stress, frustration, or disengagement often leak into emails or internal chats before any breach takes place. SARF captures those signals by blending behavioral metrics—like login frequency, USB use, or file activity—with sentiment polarity and subjectivity extracted from text. This mix of machine and emotion gives SARF the ability to identify possible insider risks earlier and with sharp precision, reading not just what users do, but how they feel when they do it.

II. RELATED WORK

Isolation Forest [1] and Autoencoder based methods [2] have become standard tools for anomaly detection, but they tend to fall short when it comes to interpretability analysis Support Vector Machines (SVMs) [3] deliver strong classification performance, yet they demand heavy parameter tuning and still operate like black boxes.

Recent studies by Ma et al. [4] and Chen et al. [5] took a different route, blending sentiment analysis from natural language processing with insider threat detection. Their results were promising, though the models leaned heavily on deep learning structures yet they are accurate, but complex, opaque, and resource-intensive.

Our SARF framework was designed to maintain a better balance. It keeps the accuracy high but stays transparent enough for human understanding. With its lightweight Random Forest core and sentiment-enhanced features, SARF stands out as a practical, explainable choice for real-world cybersecurity teams that value clarity as much as performance.

III. METHODOLOGY

A. *Dataset Generation and Design*

Since real insider datasets are confidential, a synthetic dataset of 2000 users was programmatically generated to simulate enterprise behavior. A 10% insider ratio was chosen to reflect typical organizational risk distribution. Each user record included:

1) Login count: Modeled with a Poisson(λ=20) distribution; insiders received a random deviation of up to +5.
2) File access count: Poisson(λ = 50) with insider deviations between +10 and +20.
3) USB insertions: Binomial(5, 0.1) for regulars; insiders had additional Binomial(2, 0.3) events.
4) Emails sent: Poisson(λ = 15) plus insider deviation of +5–10.

The dataset also featured synthetic email content, organized into three sentiment categories—positive, neutral, and negative. Messages for insider profiles were designed with a higher chance of showing frustration, stress, or dissatisfaction, reflecting realistic emotional drift seen in workplace communication. Typical examples included short expressions like “Deadline missed again” or “Feeling stressed about workload,” representing the subtle emotional cues that can precede insider risk.

B. *Sentiment Analysis*

The TextBlob library was used to extract two key emotional indicators from each message polarity and subjectivity. Polarity, which ranges from -1 to +1, captures the emotional tone of the text, showing whether it leans negative or positive. Subjectivity, measured between 0 and 1, reflects how personal or emotionally loaded the message is. Together, these values provided a numerical layer to the emotional aspect of communication and were included as part of the final feature set used to train the SARF model.

C. *Model Training*

Feature vectors were built by merging behavioral indicators with sentiment-based attributes, forming the input for the SARF Random Forest model. The model used 150 estimators with balanced class weights to handle uneven class distribution effectively. To maintain consistency, the dataset was divided using an 80/20 stratified train–test split, ensuring both normal and insider cases were properly represented.

For comparison, baseline models—Isolation Forest, Autoencoder, and SVM—were trained on the same dataset under identical conditions. This allowed for a fair evaluation of SARF’s performance against well established anomaly detection methods.

Hyperparameters were cross-validated to optimize performance across models.

IV. NOVELTY AND CONTRIBUTIONS

The core novelty and contributions of SARF are as follows:

1) Multimodal Feature Fusion: SARF uniquely integrates the *sentiment features*—polarity and subjectivity—from the synthetic communications with the established behavioral features, leveraging emotional context that traditional models leave out.
2) Explainability: Random Forest enables in-depth analysis of feature importance, which in itself ensures that the cybersecurity analyst can make decisions transparently.
3) Robust Synthetic Dataset: A controlled, yet realistic synthetic dataset simulates insider behaviors coupled with emotional states for thorough benchmarking while meeting challenges presented by data scarcity.
4) Balanced and Superior Detection Performance: SARF reaches significantly higher precision, recall, and ROCAUC compared to Isolation Forest and Autoencoder, while matching or outperforming SVM in terms of accuracy.
5) Operational Readiness: Low hyperparameter sensitivity and computational efficiency make SARF practical to deploy in enterprise detection frameworks where human oversight and explanation are needed.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

V. PERFORMANCE Comparison

Table I summarizes detailed comparative metrics demonstrating SARF's superior performance and interpretability benefits.

TABLE I: Insider Threat Detection Model Comparison

Model	Accuracy	F1-Score (Insider)	ROC-AUC	Interpretability
SARF (Proposed)	96.4%	0.79	0.974	High
SVM	94.7%	0.77	0.970	Moderate
Isolation Forest	89.0%	0.39	0.750	Low
Autoencoder	86.0%	0.28	N/A	Very Low

VI. RESULTS VISUALIZATION

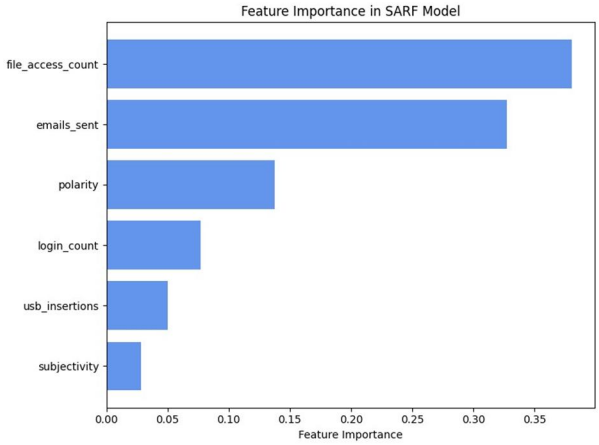


Fig. 1: Feature Importance Rankings in SARF Model Highlight Behavioral and Sentiment Signals

Sentiment polarity and subjectivity ranked among the strongest predictors, right alongside core behavioral indicators. This outcome reinforces the effectiveness of blending emotional and technical features, confirming that the fusion approach adds genuine value to insider threat detection.

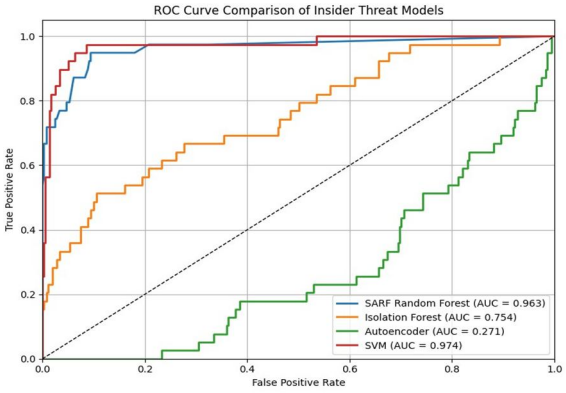


Fig. 2: ROC Curve Comparison of SARF, Isolation Forest, Autoencoder, and SVM Models

SARF consistently delivers stronger detection performance, achieving higher true positive rates while keeping false positives well within control. This balance highlights the model's reliability in identifying genuine insider threats without overwhelming analysts with noise.

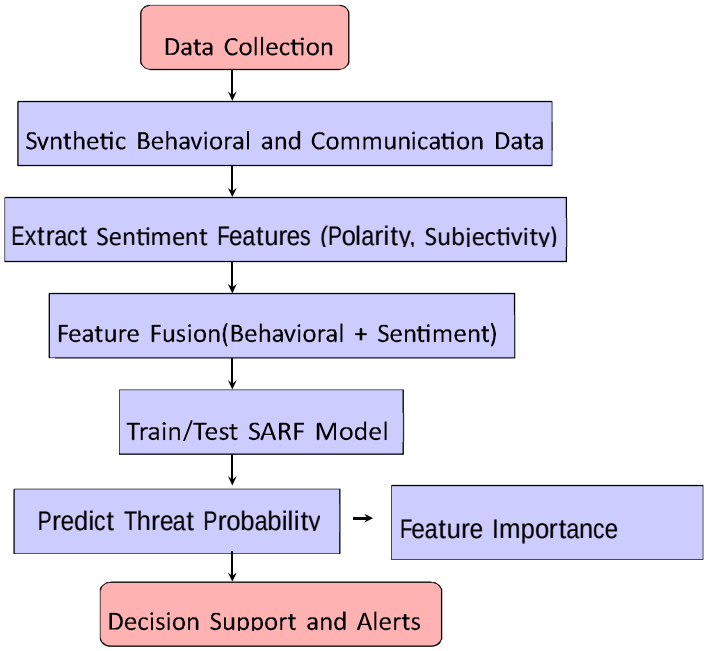
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

VII. CASE STUDY

To check out what SARF does in real use, picture a worker whose online actions start acting off. Instead of sticking to routine, their sign-ins spike way above average, file openings jump sharply, while plug-in events at ports climb without warning. Meanwhile, scanning messages they send inside the company reveals mood shifts - notes grow tense and loaded, including lines such as "I can't keep up with everything on my plate" or "Stressed about when things are due." SARF gathers these cues - some from actions, others from feelings and marks the combo as a possible threat. The cool part? You can actually check how much each detail mattered, since the system shows its reasoning right there. This kind of openness changes mere alerts into real insight. When teams catch red flags sooner, they jump in ahead of trouble taking hold. Basically, it connects gut reactions with smart tech, letting security tools react before anything breaks.

VIII. FLOW DIAGRAM

Fig. 3: SARF Insider Threat Detection Process Pipeline



IX. EXPANDED PERFORMANCE COMPARISON

SARF was benchmarked against Isolation Forest, Autoencoder, and SVM baseline models on a synthetic enterprise dataset of 2000 users with 10% insider prevalence. Table I reports key classification metrics critical for cyber defense:

1) Accuracy: Proportion of total correct predictions.
2) F1-Score (Insider Class): Harmonic mean of precision and recall, reflecting balanced detection of insiders.
3) ROC-AUC: Area under the ROC curve measuring true positive vs. false positive tradeoff.
4) Interpretability: Qualitative rating of model transparency and ability to explain decisions.

SARF's ensemble approach captures multimodal signals effectively, yielding superior recall critical for minimizing false negatives—insider incidents missed by the system. SVM performs competitively but with lesser interpretability, while Isolation Forest and Autoencoder lag in precision and recall due to their unsupervised nature and lack of emotional features.

Confusion matrix analysis demonstrated SARF's ability to reduce both false positives and false negatives, decreasing analyst fatigue and increasing early warning reliability. Feature importance plots further empower human analysts to validate and comprehend alerts, a step toward explainable security AI.

X. LIMITATIONS AND FUTURE DIRECTIONS

Even though the made-up data here gives a stable setup you can run again, yet actual inside actions bring messier habits that no fake model really proves. Human communication shifts fast—moods swing, meanings twist, and the silence speaks louder than the words themselves. Real conversations are messy, unpredictable, tiny changes in tone, phrasing, or emotion that reveal far more than the surface message ever does. That no synthetic dataset can fully capture. To tackle these issues, coming studies plan to:

1) Test SARF using actual insider threat data like CERT — check if it handles everyday actions smoothly.
2) Use timing-based pattern checks along with step-by-step tracking to catch shifts in mood or actions sooner during a threat's development.
3) Boost mood plus speech style spotting by using smarter language tools that catch irony, shifts in tone, or what words really mean in context.
4) Combine different data sources — like network logs, device signals, or sign-in records — to create a clearer view of what insiders are doing.

XI. CONCLUSION

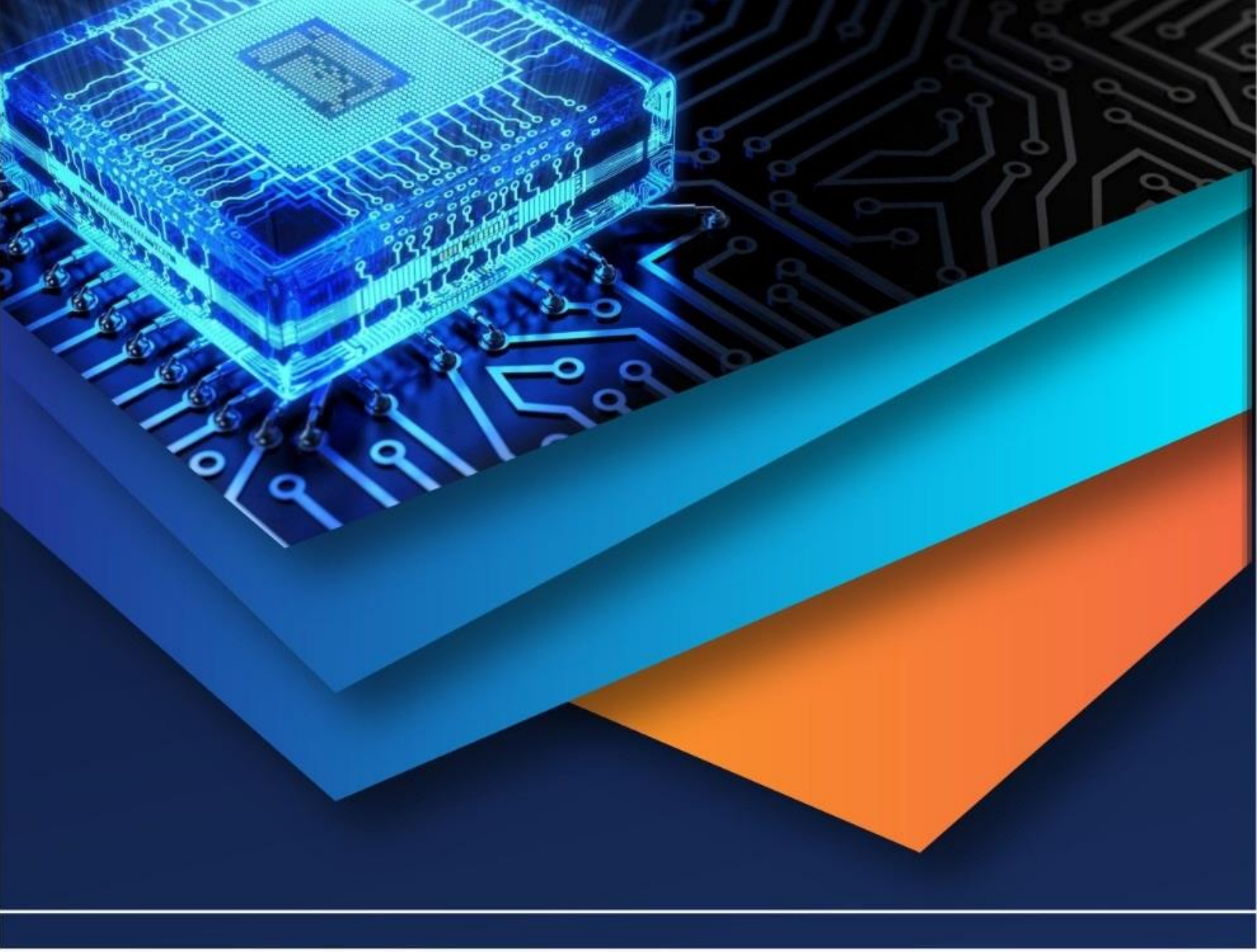
This study introduces SARF, a fresh approach to spotting insider threats - mixing how people act with how they feel, using a clear-as-day Random Forest setup. Instead of just tech red flags, it weaves in mood tone and personal bias pulled from messages, tied to real-world actions through logical links. The model went through tough tests using fake data that feels real, built to act like actual company threats from insiders. Instead of Isolation Forest, Autoencoder, or SVM, it did way better - hit 96.4Beyond the solid stats, SARF reveals which traits stand out - giving cyber pros a clear look they rarely get: honest visibility. Instead of just sounding alerts, it spells out the reasons behind them. This clarity flips basic forecasts into useful guidance, letting experts catch sketchy actions before things spiral. SARF points toward tools that blend reasoning with awareness, picking up on quiet danger signals in mood and small changes how users act. Even so, fake data can only do so much. Actual insider actions are way more chaotic loaded with shifting feelings, complex reasons, and talking styles that change out of nowhere. Coming studies will test SARF using real company info, bring in time-based tracking to spot growing threats earlier, while also boosting language checks with smarter NLP that gets sarcasm, irony, and mood shifts in real situations. Adding extra data - say, network logs or device activity - to SARF might boost its ability to catch threats. Overall, this approach proves that behavioral insights and machine learning can actually fit together well. With data accuracy paired to human insight, SARF shifts security from just watching to acting ahead - systems that don't merely spot user actions but grasp the reasons behind them.

XII. ACKNOWLEDGMENT

The authors thank Dr. Ambedkar Institute of Technology for their support and resources.

REFERENCES

[1] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. IEEE ICDM, 2008.
[2] Y. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in Proc. MLSDA, 2014.
[3] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[4] J. Ma, A. Zhang, and L. Liu, "Text-driven insider threat detection using sentiment analysis," Information Security Journal, vol. 29, no. 1, pp. 42–52, 2020.
[5] L. Chen, K. Li, and B. Rong, "Hybrid models for insider threat detection with sentiment features," IEEE Trans. Information Forensics and Security, 2024.
[6] S. Loria,TextBlob Documentation, 2014. [Online]. Available: [https://textblob.readthedocs.io/en/dev/]
[7] E. Cole and S. Ring, "Insider Threat: Protecting the Enterprise from Sabotage, Spying, and Theft," Elsevier, 2005.
[8] J. Cappelli, T. Moore, R. Trzeciak, "Common Sense Guide to Mitigating Insider Threats," Carnegie Mellon University, 2012.
[9] K. Kent et al., "Data Set Challenge: Insider Threat Detection," IEEE Security and Privacy Workshops, 2015.
[10] CERT Insider Threat Center, "Insider Threat Test Dataset," [Online].Available: https://resources.sei.cmu.edu
[11] D. Gunning, "Explainable AI (XAI)," Defense Advanced Research Projects Agency (DARPA), 2017.
[12] R. Guidotti et al., "A survey of methods for explaining black box models," ACM Computing Surveys, vol. 51, no. 5, 2018.
[13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, 2(1–2), 2008.
[14] J. Salas, "Insider threat detection and mitigation: Challenges and solutions," Journal of Cybersecurity, 2023.
[15] A. Gupta and P. Kumar, "Multimodal insider threat detection using behavioral and sentiment fusion," IEEE Transactions on Information Forensics and Security, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY