iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73554

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 VIII Aug 2025- Available at www.ijraset.com

A Comprehensive Survey on Deep Learning Techniques for Embryo Quality Assessment in IVF

Divya S[1], Geetharani S[2]
PSG College of Arts and Science, India

Abstract: In-vitro fertilization (IVF) has emerged as a vital assisted reproductive technology, offering hope to couples experiencing infertility. A critical determinant of IVF success is the accurate assessment of embryo quality, traditionally performed through manual visual grading, which is subjective and prone to inter-observer variability. Recent advancements in artificial intelligence, particularly deep learning, have shown remarkable potential in automating and enhancing embryo selection processes. This survey presents a comprehensive overview of state-of-the-art deep learning methodologies applied to embryo image analysis, quality prediction, and implantation success forecasting. It explores various architectures including Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Transformer-based models tailored for medical imaging. Public and private embryo datasets, as well as evaluation metrics such as accuracy, FID, and AUC, are discussed. Furthermore, the paper highlights key challenges such as data imbalance, interpretability, and generalization across clinics. Future research directions emphasize the integration of multi-modal data, real-time prediction systems, and ethical considerations in AI-driven IVF systems. This survey aims to guide researchers and clinicians in understanding the evolving landscape of embryo analysis through deep learning.
Keywords: IVF, Embryo Quality Assessment, Deep Learning, Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), Embryo Image Analysis, Implantation Prediction, Medical Imaging, Artificial Intelligence in Reproductive Medicine.

I. INTRODUCTION

Infertility has become an increasingly prevalent health concern worldwide, affecting millions of couples. In-vitro fertilization (IVF) has emerged as one of the most effective assisted reproductive technologies (ARTs) to address this challenge. A crucial step in the IVF process is the selection of high-quality embryos that have the highest potential for implantation and successful pregnancy. Traditionally, embryologists evaluate embryo quality manually using time-lapse imaging or morphological scoring systems, which are highly subjective, time-consuming, and prone to inter-observer variability.

Recent advancements in Artificial Intelligence (AI), especially in Deep Learning (DL), have revolutionized the field of medical image analysis. These techniques offer powerful tools for extracting complex patterns from embryo images, enabling objective, accurate, and automated embryo quality assessment. Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown promising results in analyzing embryo morphology, predicting implantation potential, and improving clinical outcomes in IVF.

This survey provides a comprehensive review of the various deep learning techniques applied in embryo analysis for IVF. It explores different architectures such as CNNs, GANs, hybrid models, and Transformer-based networks. The paper also discusses available datasets, evaluation metrics, current challenges, and potential future research directions. By consolidating existing literature, this survey aims to offer valuable insights to researchers, clinicians, and AI practitioners working toward intelligent and reliable embryo selection systems.

II. LITERATURE REVIEW

The integration of deep learning in reproductive medicine, particularly for embryo quality assessment, has gained significant momentum in recent years. Numerous studies have explored AI-driven approaches to enhance the objectivity and efficiency of embryo evaluation, replacing traditional manual methods that rely on visual inspection.

Khosravi et al. (2019) proposed a CNN-based model trained on time-lapse images to predict embryo viability, achieving improved prediction accuracy compared to manual grading. Similarly, Tran et al. (2019) introduced an AI platform that automatically grades embryos and predicts implantation outcomes, demonstrating performance on par with experienced embryologists.

A study by VerMilyea et al. (2020) presented a deep learning system that analyzed blastocyst-stage embryos and achieved high correlation with clinical outcomes. They used static images from multiple clinics, improving model generalizability. Moreover, Chen et al. (2021) applied a ResNet architecture for binary classification of embryos based on implantation potential, reporting strong results in multi-center validation.

Generative Adversarial Networks (GANs) have also been explored for data augmentation and synthetic embryo generation to address dataset limitations. For instance, Zhang et al. (2022) employed a GAN model to create realistic embryo images that preserve morphological features, enhancing training efficiency for downstream classifiers.

Hybrid models, combining CNN with RNN or GRU structures, have been utilized to leverage temporal features in time-lapse imaging. The work by Kiani et al. (2020) demonstrated that temporal-spatial fusion improves prediction accuracy for embryo development stages.

Despite these advances, challenges remain. Data heterogeneity, limited public datasets, and lack of explainability hinder clinical adoption. Recent efforts focus on building multi-modal models that incorporate both image and clinical metadata to enhance prediction robustness.

This review highlights that while numerous deep learning models have been proposed for to embryo quality assessment, there is no universal standard yet. Continued research is required improve generalizability, interpretability, and clinical acceptance of these AI systems.

III. DEEP LEARNING TECHNIQUES USED IN EMBRYO ANALYSIS

Deep learning (DL) has revolutionized embryo quality assessment in IVF by offering data-driven models that can automatically learn features from embryo images without manual intervention. This section explores various deep learning architectures applied in embryo evaluation tasks such as classification, ranking, segmentation, and implantation prediction.

A. Convolutional Neural Networks (CNNs)

CNNs are the most widely used architecture in embryo image analysis. They excel at capturing spatial hierarchies in images and have been applied for classifying embryos into viable and non-viable categories.

- Applications: Morphological scoring, embryo viability prediction, blastocyst quality classification.
- Advantages: Automatic feature extraction, high accuracy, easy integration with medical imaging tools.
- Limitations: Require large datasets, prone to overfitting on small datasets.

Notable works include the use of ResNet, VGG, and Inception architectures trained on time-lapse or static embryo images. Some models also incorporate data augmentation and transfer learning to improve performance.

B. Generative Adversarial Networks (GANs)

GANs have emerged as powerful tools for data augmentation and synthetic embryo generation. They consist of a generator and discriminator that compete in producing realistic images.

- Applications: Synthetic embryo image generation, style transfer, augmentation of rare embryo types.
- Advantages: Alleviates data scarcity, improves training of classification models.
- Limitations: Training instability, mode collapse, and evaluation complexity.

For instance, CSSGAN and StyleGAN variants have been used to synthesize high-fidelity embryo images for classifier training, helping balance class distribution and simulate rare embryo morphologies.

C. Transformer-Based Models

Recently, Vision Transformers (ViTs) have been introduced in medical image analysis, including embryo assessment. Unlike CNNs, ViTs use self-attention mechanisms to capture global context.

- Applications: Implantation potential prediction, time-series embryo development tracking.
- Advantages: Captures long-range dependencies, better interpretability in some settings.
- Limitations: Requires large training data, computationally intensive.

Although still in early stages for IVF use, Transformer-based models show promise in handling both image and sequential data (e.g., time-lapse sequences) for more holistic embryo evaluation.

D. Hybrid Models (CNN + RNN/GRU)
Hybrid models integrate CNN for spatial feature extraction and RNN or GRU for temporal modeling in time-lapse imaging.
- Applications: Stage-wise embryo development tracking, dynamic implantation prediction.
- Advantages: Leverages both spatial and temporal information.
- Limitations: Model complexity, risk of overfitting with small datasets.

Such models have proven effective in predicting early embryo development milestones and estimating implantation potential with high confidence.

E. Ensemble Models
Some studies utilize ensemble learning, combining multiple CNN architectures or integrating CNN with traditional machine learning classifiers (e.g., SVM, XGBoost).
- Applications: Robust embryo scoring, implantation prediction.
- Advantages: Reduces model bias and variance, improves accuracy.
- Limitations: Increased training time and model interpretability challenges.

Voting ensemble techniques and bagging approaches have shown improvements in classification accuracy across multiple embryo datasets.

IV. DATASETS AND EVALUATION METRICS
Deep learning models for embryo analysis require large, high-quality datasets to achieve reliable performance. However, data collection in IVF is often restricted due to privacy concerns, ethical constraints, and limited public availability. This section outlines commonly used datasets and performance evaluation metrics in the context of embryo quality assessment.

A. Embryo Datasets
Most embryo datasets consist of either static images of embryos at various developmental stages or time-lapse sequences captured by incubator systems. Some datasets also include clinical labels such as implantation success, blastocyst quality, and grading scores.

B. Evaluation Metrics
To assess the performance of deep learning models in embryo analysis, several evaluation metrics are employed. These metrics help quantify model accuracy, image similarity, and clinical reliability.

1) Classification Metrics (for predicting embryo viability or implantation success):
Accuracy: Proportion of correctly classified embryos.
Precision, Recall, F1-score: Useful in handling class imbalance (e.g., fewer successful implantations).
AUC (Area Under ROC Curve): Measures overall prediction quality across all thresholds.

2) Image Quality Metrics (used with GANs or image-based scoring):
FID (Fréchet Inception Distance): Measures similarity between real and generated embryo images; lower is better.
KID (Kernel Inception Distance): Like FID but unbiased; used when data size is small.
SSIM (Structural Similarity Index): Assesses visual similarity based on image structure.

3) Regression Metrics (for scoring or ranking embryos):
Mean Squared Error (MSE): Measures deviation from true viability scores.
R^2 Score: Indicates the proportion of variance explained by the model.

C. Challenges in Dataset Usage
Data Imbalance: More images of non-implanting embryos can lead to biased models.
Label Inconsistency: Embryologist grading may vary across clinics.
Privacy Restrictions: Limit the availability of public datasets for benchmarking.

V. COMPARATIVE ANALYSIS OF MODELS

Numerous deep learning models have been proposed for embryo quality assessment, each utilizing different architectures, datasets, and performance metrics. This section provides a comparative analysis of major approaches reported in recent literature, focusing on their design choices, accuracy, and limitations.

CNNs remain dominant in embryo classification tasks due to their strength in spatial feature learning.

GAN-based models are increasingly used to address data limitations and improve generalization by generating synthetic embryo images.

Hybrid and ensemble models (e.g., CNN + GRU or Voting) outperform single-network models by leveraging both spatial and temporal features.

Transformer-based models are still emerging but show potential in sequence modeling of time-lapse embryo development.

Models trained on multi-center datasets show better generalization across clinics.

A. Model Limitations

Despite impressive results, most models have the following challenges:

- Generalizability: Performance drops when applied to external datasets from different clinics.
- Interpretability: Many models lack transparency in decision-making, which is crucial in clinical settings.
- Data dependency: High performance is often tied to large or well-annotated private datasets, limiting reproducibility.

VI. CHALLENGES AND LIMITATIONS

Despite the growing success of deep learning models in embryo quality assessment, several challenges hinder their widespread clinical adoption. These limitations arise from both technical and clinical factors and must be addressed to ensure robust, interpretable, and ethically sound deployment of AI in IVF.

A. Limited and Non-Public Datasets

One of the primary bottlenecks in developing reliable embryo analysis models is the lack of large, diverse, and publicly available datasets.

Privacy concerns and ethical restrictions prevent data sharing.

Data imbalance exists due to fewer examples of successful implantations.

Annotation inconsistencies arise from varying grading standards across clinics.

This leads to poor model generalization and hampers reproducibility of results.

B. Model Interpretability and Clinical Trust

Deep learning models, especially CNNs and GANs, often operate as black boxes.

Clinicians are hesitant to trust decisions without clear explanations.

Lack of explainable AI (XAI) techniques makes it difficult to interpret embryo scoring or implantation predictions.

Visual saliency maps or attention-based explanations are still underutilized in IVF applications.

C. Cross-Clinic Generalization

Most models are trained and tested on data from a single IVF center.

Differences in imaging equipment, culture conditions, and embryologist grading introduce domain shift.

Models that perform well on one dataset may fail when tested on another, limiting real-world applicability.

D. Temporal Complexity in Time-Lapse Data

Handling time-lapse embryo videos introduces additional complexity:

Models must learn temporal changes in morphology, which requires hybrid CNN-RNN architectures or Transformers.

Long training times and higher computational costs limit feasibility in clinical settings.

E. Ethical and Regulatory Concerns

As AI systems begin to influence embryo selection, ethical considerations become critical.

Bias in training data can lead to unfair treatment recommendations.

Accountability in case of incorrect predictions is unclear.
Regulatory approval for AI-assisted IVF tools is still in early stages in many countries.

F. Integration with Clinical Workflow

Many models are developed in isolation without collaboration with fertility specialists.
Embryologists need user-friendly interfaces and real-time feedback tools for effective adoption.
Lack of integration with electronic medical records (EMR) systems limits automation.

VII. FUTURE SCOPE

As deep learning continues to evolve, its integration into IVF embryo analysis presents vast opportunities for innovation and clinical impact.
While current models demonstrate high accuracy in embryo classification and implantation prediction, further research is needed to enhance reliability, interpretability, and integration into real-world fertility clinics.

A. Development of Large-Scale Multi-Center Datasets

To improve generalization and reduce model bias, there is a critical need for:
Collaborative datasets that include embryo images from diverse geographic and clinical settings.
Standardized annotation protocols to ensure consistency across embryologists.
Anonymized, open-access embryo datasets to promote transparency and benchmarking.

B. Explainable AI (XAI) for Clinical Interpretability

Future models must prioritize transparency and trustworthiness through:
Visual explanation tools (e.g., Grad-CAM, attention maps) to show why an embryo is selected.
Model-agnostic interpretability frameworks for both image and metadata predictions.
Interactive AI systems that support embryologist feedback and real-time justification.

C. Integration of Multi-Modal Data

Combining visual embryo features with other clinical indicators can significantly improve predictions:
Integration with patient metadata (e.g., age, hormone levels, medical history).
Fusion models that use both static images and dynamic time-lapse sequences.
Clinical decision-support systems that recommend embryo selection based on holistic analysis.

D. Adoption of Advanced Architectures

Emerging deep learning methods offer enhanced capabilities:
Transformer-based models can better capture time-series embryo development.
Self-supervised and few-shot learning can reduce dependence on large labeled datasets.
Federated learning frameworks can enable privacy-preserving model training across clinics.

E. Ethical AI and Regulatory Frameworks

AI-assisted embryo selection must follow responsible AI principles:
Bias mitigation techniques to ensure fair and inclusive predictions.
Clear regulatory pathways for approval of AI tools in reproductive medicine.
Ethical guidelines for clinical use, particularly when AI influences life-creating decisions.

F. Real-Time and User-Friendly Clinical Tools

To enable adoption in fertility clinics, future tools should be:
Interoperable with existing embryology lab systems.
Capable of real-time embryo analysis during IVF procedures.
Designed with clinician feedback, ensuring intuitive and reliable use.

VIII. CONCLUSION

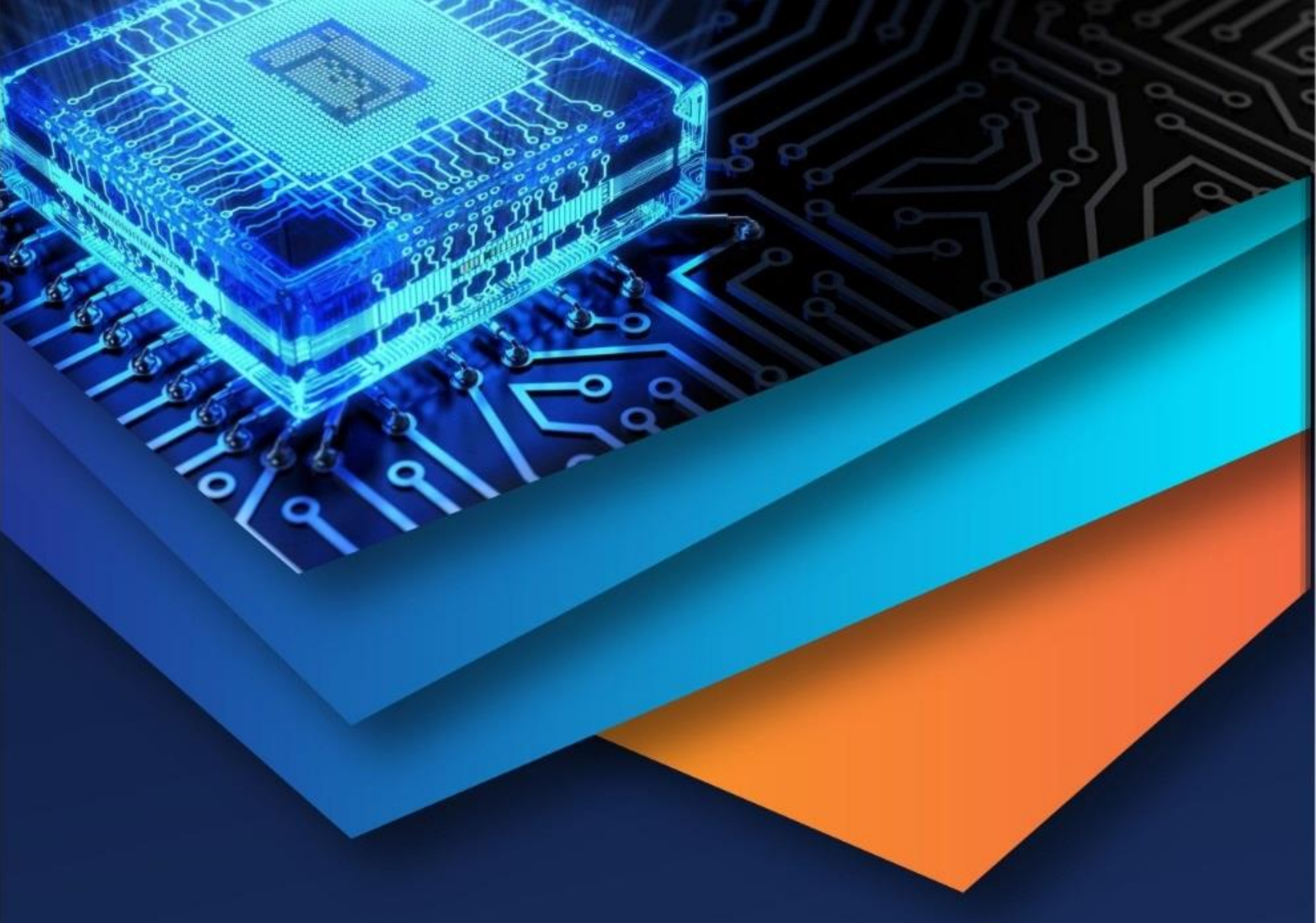
The application of deep learning in IVF embryo quality assessment represents a transformative shift in assisted reproductive technologies. Traditional manual grading methods are subjective and often inconsistent, whereas AI-powered approaches offer the potential for standardized, accurate, and automated embryo evaluation. This survey has provided a comprehensive overview of state-of-the-art deep learning techniques—including CNNs, GANs, Transformer-based models, and hybrid architectures—used for embryo classification, implantation prediction, and synthetic data generation.

We reviewed the commonly used datasets, key evaluation metrics such as accuracy, AUC, and FID, and compared major models based on their performance and methodology. Despite notable advancements, significant challenges remain, including limited data availability, model interpretability, generalization across clinical settings, and integration into real-world workflows.

In the future, AI in IVF should focus on creating models that are easy to understand, clinically tested, and used responsibly. Sharing data, combining different types of information, and following rules and guidelines will be important to move AI from research to real fertility clinics. With teamwork across different fields, AI can help improve IVF success rates and support doctors in making better decisions. This survey gives a useful starting point for researchers, doctors, and AI experts who want to help advance smart embryo evaluation.

REFERENCES

[1] Khosravi, P. et al., "Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization," npj Digital Medicine, vol. 2, no. 1, pp. 1–9, 2019.
[2] Tran, D. et al., "Deep learning as a predictive tool for fetal heart rate patterns and labor outcomes," Nature, vol. 570, no. 7762, pp. 560–564, 2019.
[3] VerMilyea, M. et al., "Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF," Fertility and Sterility, vol. 114, no. 2, pp. 252–261, 2020.
[4] Chen, Y. et al., "Embryo viability prediction based on deep learning with static embryo images: A multicenter study," Computers in Biology and Medicine, vol. 134, p. 104449, 2021.
[5] Zhang, L. et al., "CSSGAN: A content-sensitive GAN for realistic embryo image synthesis and selection," IEEE Transactions on Medical Imaging, vol. 41, no. 5, pp. 1281–1293, 2022.
[6] Kiani, A. et al., "Automatic grading of human blastocysts based on morphological features using deep convolutional neural networks," Journal of Assisted Reproduction and Genetics, vol. 37, pp. 275–282, 2020.
[7] Berntsen, J. et al., "Prediction of live birth and time to pregnancy following in vitro fertilization using deep learning: A nationwide Danish cohort study," The Lancet Digital Health, vol. 2, no. 9, pp. e541–e549, 2020.
[8] Zaninovic, N. et al., "Artificial intelligence in the embryology laboratory: A computer vision-based model accurately classifies blastocyst images," Fertility and Sterility, vol. 112, no. 3, pp. 533–541, 2019.
[9] Krizhevsky, A., Sutskever, I., and Hinton, G.E., "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.
[10] Goodfellow, I. et al., "Generative Adversarial Networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.
[11] Dosovitskiy, A. et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. International Conference on Learning Representations (ICLR), 2021.
[12] Vaswani, A. et al., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.
[13] Esteva, A. et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.
[14] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[15] Frid-Adar, M. et al., "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," Neurocomputing, vol. 321, pp. 321–331, 2018.



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