iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: X Month of publication: October 2025

DOI: https://doi.org/10.22214/ijraset.2025.74428

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 X Oct 2025- Available at www.ijraset.com

Deep Learning Comparison Guide on Structured and Unstructured Datasets with Rare Case

Jitendra Kumar Das[1], Sudhir Panigrahy[2]
[1]Kalam Institute of technology, Berhampur
[2]NIDT University

Abstract: Deep learning has emerged as a powerful tool for analyzing both structured and unstructured datasets across diverse domains. However, its performance and suitability vary significantly depending on the data type, representation, and underlying task complexity. This paper presents a comparative study of deep learning approaches applied to structured and unstructured datasets, with special emphasis on rare-case scenarios such as imbalanced data, limited samples, and noisy environments. We review key models, architectures, and training techniques, highlighting their advantages and limitations. Experimental evidence and case references suggest that while structured datasets benefit from tabular-specific models and feature engineering, unstructured datasets rely on advanced representation learning using convolutional and transformer-based architectures. Rare-case handling techniques, including data augmentation, transfer learning, and generative modeling, are also discussed. This comparative guide aims to assist researchers and practitioners in selecting suitable deep learning strategies for specific dataset types and challenges.
Keywords: Deep Learning, Structured Data, Unstructured Data, Rare Case Handling, Comparison Guide, Neural Networks

I. INTRODUCTION

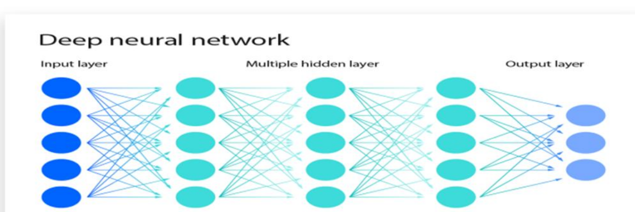
Deep learning (DL) techniques have revolutionized artificial intelligence, enabling breakthroughs in areas such as computer vision, natural language processing (NLP), and medical diagnosis. Datasets used in DL are broadly categorized into **structured** (e.g., tabular, relational databases, transactional records) and unstructured (e.g., images, audio, text, videos).
Structured data has a fixed schema and fits neatly into rows and columns, such as names and phone numbers. Unstructured data has no fixed schema and can have a more complex format, such as audio files and web pages.
A further challenge arises when dealing with rare cases, such as:
1) Highly imbalanced datasets (e.g., fraud detection, rare disease diagnosis).
2) Limited samples available for training (few-shot learning).
3) Noisy or incomplete data.
This paper provides a systematic comparison guide on applying deep learning to structured and unstructured datasets, and explores strategies to handle rare cases effectively.

II. BACKGROUND AND RELATED WORK

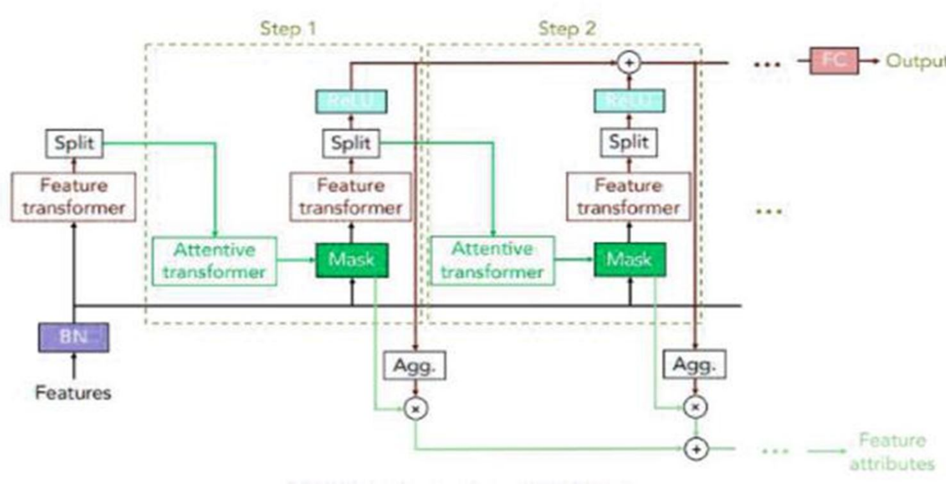
A. Structured Datasets

Traditionally analyzed using tree-based models (e.g., XGBoost, Random Forests). Deep learning models like Deep Neural Networks (DNNs), TabNet, and entity embeddings have recently been introduced for structured/tabular learning.

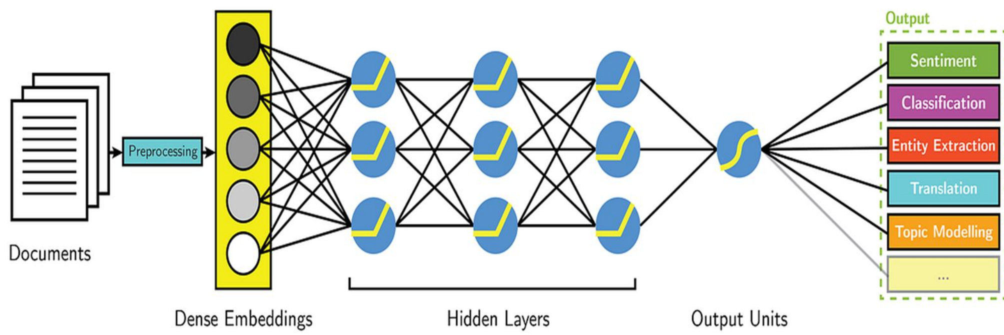
- Structured data has a fixed schema and fits neatly into rows and columns, such as names and phone numbers.
- Structured data include bot quantitative and qualitative data such as (dates, names, address etc.)



(Fig:1 DNN)



(Fig:2 Tabnet

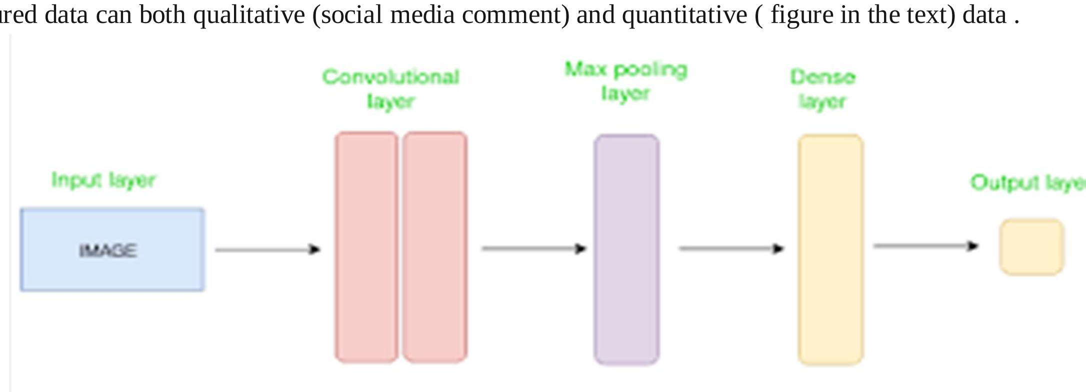


(Fig:3 Embedded dataset)

B. Unstructured Datasets

CNNs, RNNs, and Transformers dominate applications in computer vision, NLP, and speech. Representation learning is a key strength of DL for unstructured data.

- Unstructured data has no fixed schema and can have a more complex format, such as audio files and web pages.
- Unstructured data can both qualitative (social media comment) and quantitative (figure in the text) data .

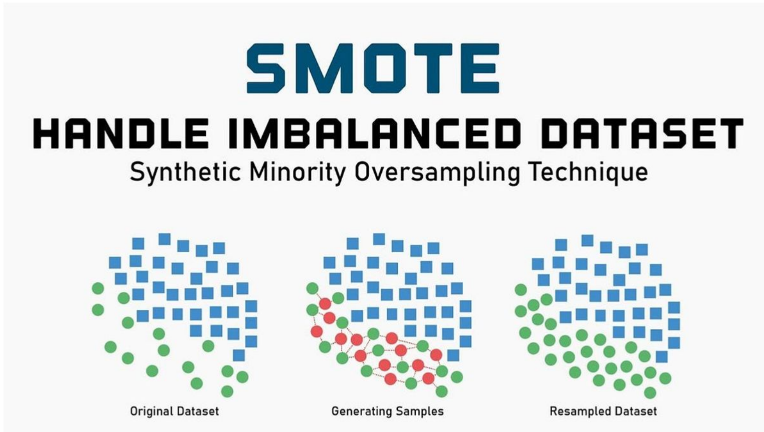


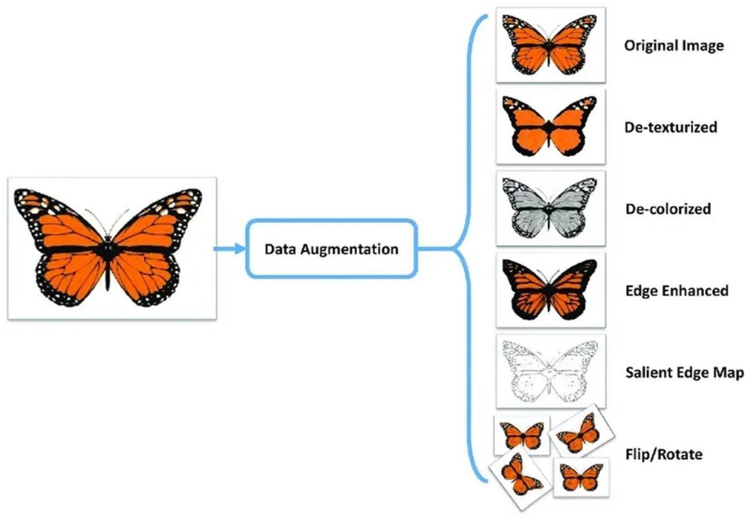
C. Rare Case Handling

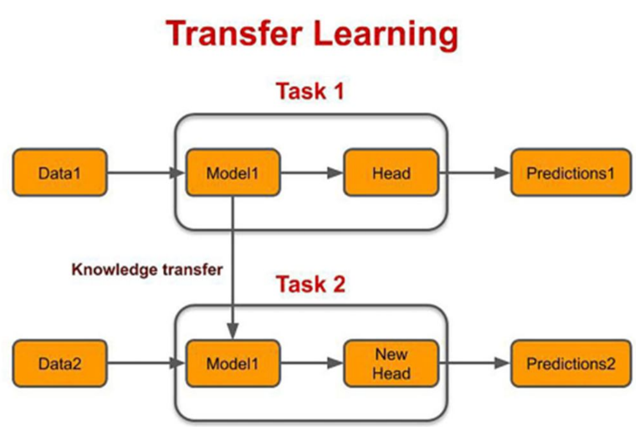
Techniques such as SMOTE (Synthetic Minority Over-sampling), data augmentation, transfer learning, and generative adversarial networks (GANs) are used to overcome data imbalance and scarcity.

- smote is a preprocessing is a techinqe for addressing is imbalance or very small miniority class . it address the issue of rare case by synthesizing them , but it can sometime produce low quality

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 X Oct 2025- Available at www.ijraset.com







Several studies have compared DL with classical machine learning, but few provide a comprehensive structured vs. unstructured comparison with rare-case focus. This paper attempts to bridge that gap.

III. COMPARISON GUIDE

A. Structured Datasets

1) Typical format: relational tables, rows & columns.
2) Example domains: finance, healthcare records, sales prediction.
3) Challenges: feature scaling, categorical encoding, sparsity.

4) DL Approaches:
   - Deep Neural Networks (MLPs).
   - TabNet (attention-based).
   - Entity embeddings for categorical features.

B. *Unstructured Datasets*

1) Typical format: text, images, audio, video.
2) Example domains: NLP (chatbots, summarization), CV (object detection), Speech (recognition).
3) Challenges: high dimensionality, context preservation, semantic meaning.
4) DL Approaches:
   - CNNs for images.
   - RNNs/LSTMs/Transformers for text & sequential data.
   - Vision Transformers (ViTs) for large-scale vision tasks.

C. *Rare Case Scenarios*

1) Class Imbalance: Fraud detection, rare disease cases.
2) Limited Data: Few-shot learning, transfer learning, one-shot classification.
3) Noisy Data: Sensor readings, social media text.
4) Techniques:
   - Oversampling/Undersampling.
   - Data augmentation (image flips, synonym replacement).
   - Transfer learning with pretrained models.
   - GANs for synthetic data generation.

IV. EXPERIMENTAL DISCUSSION (OPTIONAL if you run tests)

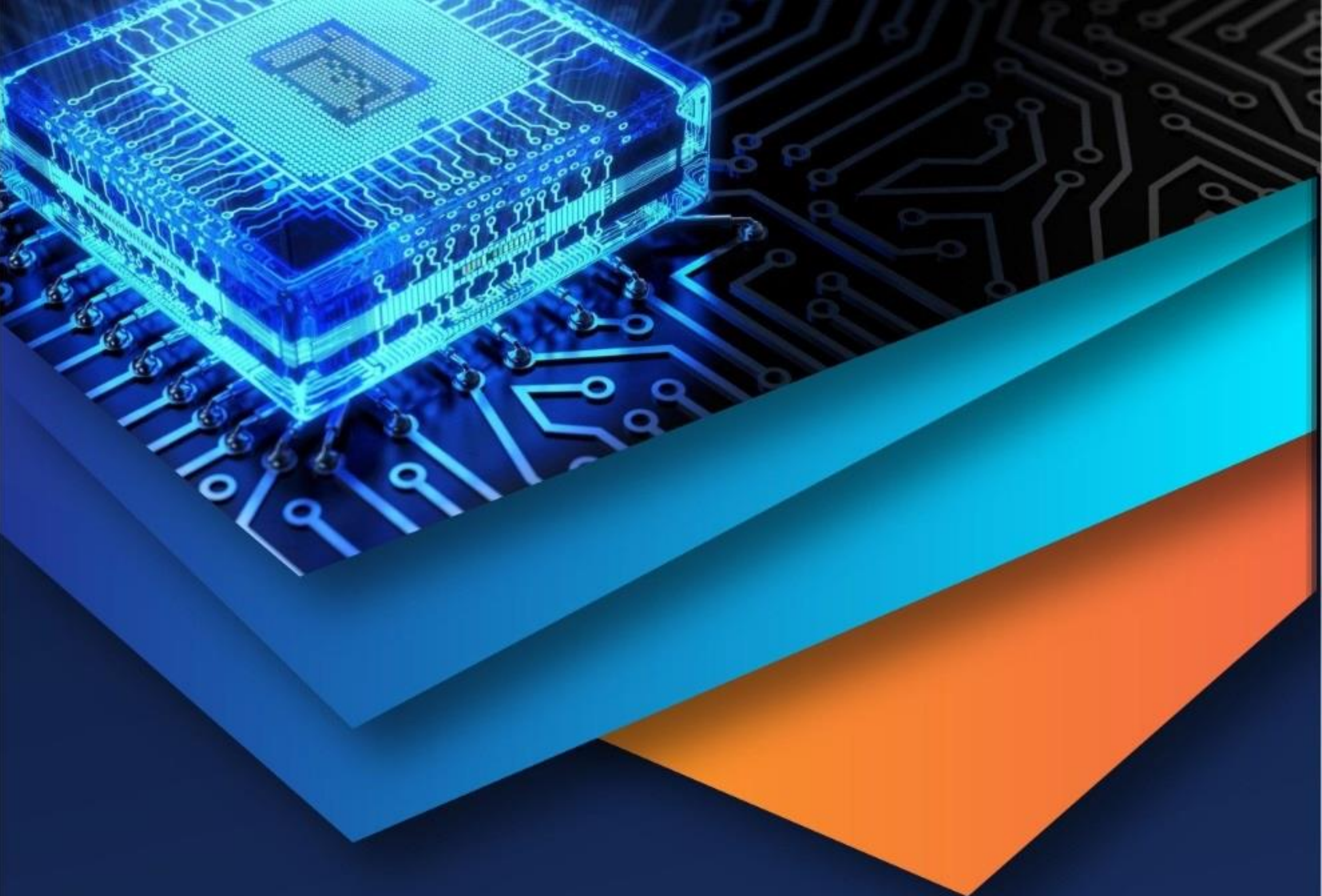
1) Compare accuracy, F1-score, ROC-AUC of DL models on structured vs. unstructured datasets.
2) Demonstrate a rare-case (e.g., imbalanced dataset like credit card fraud).
3) Highlight performance gaps and best practices.

V. CONCLUSION

This paper provides a comparative guide to deep learning techniques across structured and unstructured datasets, with a focus on rare-case challenges. Structured datasets often require specialized architectures such as TabNet or embeddings, while unstructured datasets thrive with CNNs and Transformers. Rare cases demand data-centric solutions like augmentation, GANs, and transfer learning. Future work may include benchmarking across unified datasets and developing hybrid models that combine structured and unstructured representations.

REFERENCES

(IEEE Style Samples)
[1] Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[2] A. Vaswani et al., “Attention is all you need,” Advances in Neural Information Processing Systems (NeurIPS), 2017.
[3] S. Arik and T. Pfister, “TabNet: Attentive interpretable tabular learning,” in Proc. AAAI, 2021.
[4] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” Journal of Big Data, vol. 6, no. 1, 2019.
[5] H. He and E. A. Garcia, “Learning from imbalanced data,” IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284, 2009



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY