J. Electrical Systems 20-3s (2024): 1082-1092

Recurrent Neural Networks for Image Captioning: A Case Study with LSTM

[1] **Dr. Shailaja Sanjay Mohite**

[2]**Dr. Suganthini. C**

[3]**Dr. Arunarani AR**

[4]**Dr. Lalitha Devi K**

[5]**Manish Sharma**

[6]**R. N. Patil**

[7]**Dr. Anurag Shrivastava**

Abstract: - This research investigates the viability of Long Short-Term Memory (LSTM) systems, a subtype of Recurrent Neural Networks (RNNs), for picture captioning. Leveraging the MS COCO dataset, the study compares the execution of LSTM-based RNNs with Vanilla RNN, Gated Recurrent Unit (GRU), consideration components, and transformer-based models. Experimental comes about to illustrate that the LSTM-based RNN shows competitive execution, accomplishing a BLEU-4 score of 0.72, a METEOR score of 0.68, and a CIDEr score of 2.1. The comparative investigation uncovers its prevalence over Vanilla RNN and GRU, highlighting its capability to capture long-range conditions inside successive picture information. Moreover, the study investigates the effect of consideration instruments and transformer designs, exhibiting their potential improvements in the context-aware caption era. The transformer-based show outflanks all other models, accomplishing a BLEU-4 score of 0.78, a METEOR score of 0.72, and a CIDEr score of 2.5. The findings give important bits of knowledge toward the creating scene of picture captioning strategy, which makes LSTM-based RNNs solid and productive approaches for capturing worldly groupings in visual substance. In achieving these, the study provides a framework for future developments in hybrid models and manufacturing processes that push boundaries of smart image perception and understanding.

Keywords: Long Short-Term Memory, Image Captioning, Attention Mechanisms, Recurrent Neural Networks, Transformer Architecture.

[1] Assistant Professor, Department of Mechatronics, RIT Islampur, (Autonomous under Shivaji University, Kolhapur)

shailaja.mohite1@rediffmail.com

[2]Department of School of Computer Science and Engineering, Vellore Institute of Technology,

Vellore Campus

*chinna246@gmail.com

[3]Department of Department of Computational Intelligence, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai

arunaraa@srmist.edu.in

[4]Department of Department of Computer science and Engineering, Sathyabama Institute of

Science and Technology, Chennai, India

lalithavelu10@gmail.com

[5]Associate Professor, Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun, Uttarakhand
manishsharma.cse@geu.ac.in

[6]Professor, Department of Mechanical Engineering, Bharati Vidyapeeth's College of Engineering Lavale Pune, India

rajendrakumar.patil@bharatividyapeeth.edu

[7]Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,

Chennai, Tamilnadu

Anuragshri76@gmail.com

I. INTRODUCTION

In recent years, the association of computer vision and character language processing has ignited revolutionary advancements in image recognition & description. A major area of research in this intersection is image captioning, a task that involves generating written descriptions for visual content. This research delves into the utilizing Recurrent Neural Network for picture inscription with detailed focus on Long Short-Term Memory architecture. The core difficulty of image captioning is created by the semantic gap between visual data and natural language descriptions. The traditional methods often struggled to capture the fine details, temporal relationships and significant specifics depicted in images. RNNs, intended to process sequential data, grown as a viable solution due their capacity in representing state dependence over time [1]. With regards to various RNN design, LSTMs have shown very outstanding capability in controlling the vanishing gradient problem that they are particularly proficient at arranging long-range dependencies of sequential information. This research is spurred by the want to upgrade the quality and coherence of created picture captions through a comprehensive investigation of LSTM-based RNNs. By examining the perplexing exchange between visual highlights and etymological structures, the study points to contribution bits of knowledge that progress the current state of picture captioning techniques [2]. The consecutive nature of LSTMs permits them to capture the advancing setting inside an image, facilitating the era of more relevantly significant and semantically wealthy captions. The importance of this research expands to different applications, counting assistive advances for the outwardly impeded, substance ordering for look motors, and human-computer interaction. In addition, a more profound understanding of the qualities and impediments of LSTM-based RNNs in picture captioning can clear the way for progressed models and techniques in related areas. Within the subsequent areas, we are going dig into the hypothetical underpinnings of RNNs, with a particular focus on LSTMs, and display a detailed technique for our case study [3]. The exploratory results and discourses that take after aim to shed light on the execution of LSTM-based RNNs within the challenging assignment of picture captioning, eventually contributing to the broader scene of shrewd picture investigation and elucidation.

II. RELATED WORK

GAAFAR et al. [15] conducted a comparative examination of the execution of profound learning classification approaches based on LSTM-RNNs for printed and picture datasets. The study dives into the flexibility of LSTM-RNNs in handling both modalities, highlighting their viability in capturing successive conditions. While the focus is on classification, the experiences gained are significant to understanding the broader capabilities of LSTM-RNNs in dealing with sequential information, laying the basis for our investigation within the context of picture captioning. In a diverse space, GAFUROV et al. [16] connected LSTM systems to progress agrarian trim acknowledgement utilizing satellite information. The ponder exhibits the appropriateness of LSTM systems in spatial information examination, emphasizing their part in capturing worldly conditions in satellite symbolism. In spite of the fact that the essential centre is on trim acknowledgement, the worldly modelling capabilities of LSTM systems adjust with their utilisation in picture captioning, where capturing transient connections between objects in a picture is vital. Guo et al. [17] presented a semantic direction arranged for video captioning, emphasizing the part of consideration instruments in improving the quality of produced captions. The consideration instrument makes a difference the show centres on particular districts of intrigued inside the video outlines, making strides in the significance and instruction of the captions. This adjusts with our investigation of consideration components in picture captioning, recognizing their commitment to the context-aware caption era. Javanmardi et al. [18] proposed an inventive approach called Caps Captioning, joining made strides in capsule systems for picture captioning. Capsule systems point to capture various levelled connections between picture components, giving a novel viewpoint in differentiation from conventional convolutional neural systems (CNNs). This ponder contributes to the differences of models investigated in picture captioning, rousing our consideration of LSTM-based RNNs near other progressed models. Kaur and Kaur [20] presented an effective profound learning-based cross-breed demonstrate for the picture caption era. Whereas the study essentially centres on the effectiveness of the proposed demonstration, it underscores the persistent investigation of hybrid structures to upgrade the execution of picture captioning frameworks. This reverberates with our approach, as we consider LSTM-based RNNs in conjunction with consideration instruments for the progressed caption era. Khosravi et al. [21] dug into the application of multivariate multi-step Long Short-Term Memory (LSTM) neural systems for concurrent stream-water variable expectation. In spite of the fact that the setting varies, the study emphasizes the viability of LSTMs in taking care of multivariate time-series information. This adjusts with our utilization of LSTMs within the successive setting of picture captioning, where capturing conditions in picture highlights over time is significant. Kwon and Kim [22] and Kwon and Lee [23] investigated security viewpoints in picture captioning models. They examined antagonistic assaults, particularly centring on restricted-area and backdoor assaults for picture captioning models. These studies highlight the defenselessness of picture captioning models to antagonistic control and emphasize the requirement for strong designs, a thought we keep in mind when evaluating the unwavering quality of LSTM-based RNNs. Li and Shao [26] proposed a setting semantic assistant network for picture captioning. This study presents relevant semantic data to improve the quality of created captions. The joining of context-specific data adjusts with our investigation of LSTM-based RNNs, where capturing setting over successive information contributes to made strides caption era. Laitsos et al. [24] upgraded mechanized profound learning applications for short-term load forecasting. Whereas the essential centre is on load forecasting, the study illustrates the generalizability of profound learning approaches to time-series information. This broader perspective adjusts with our investigation of LSTM-based RNNs, emphasizing their appropriateness for taking care of consecutive information within the setting of picture captioning. Lasheen and Barakat [25] investigated Arabic image captioning, particularly investigating the impact of text pre-processing on attention weights and BLEU-N scores. This study highlights language-specific contemplations in image captioning, an aspect that resounds with our broader investigation of the caption era utilizing LSTM-based RNNs [26]. In rundown, the related work grandstands the differing approaches analysts have taken to address challenges in image captioning. From the application of LSTM-RNNs and consideration instruments to investigating inventive structures like capsule systems, the studies give profitable bits of knowledge that advise our own investigation into the effectiveness of LSTM-based RNNs for picture captioning.

III. METHODS AND MATERIALS

1. Data:

1.1 Dataset Description:

For this study, we utilized the broadly acclaimed MS COCO (Microsoft Common Objects in Context) dataset, which gives an assorted collection of images at the side human-generated captions. The dataset comprises of over 120,000 pictures traversing 80 protest categories, making it reasonable for training and assessing picture captioning models [4].

1.2 Data Preprocessing:

Prior to preparing, the pictures were preprocessed to guarantee consistency and compatibility with the chosen algorithms. This included picture resizing to a settled determination, pixel esteem normalization and information augmentation forms that were utilized to show strength [5].

2. Algorithms:

To illuminate the issue of picture captions, four obvious calculations were picked for execution and comparable examination. These computations are diverse shapes of Recurrent Neural Networks (RNNs), centering basically on the Long Short-Term Memory LSTM structures.

2.1 Vanilla LSTM:

The Vanilla LSTM acts as a case ponder for our research. This LSTM variety comprises a cell state, an input portal, and a disregard entryway and yield entryway [6]. The conditions overseeing the Vanilla LSTM are as follows:

```
“for t in range(sequence_length):
    input_gate = sigmoid(dot(W_ii, x_t) + dot(W_hi,
h_{t-1}) + b_ii + b_hi)
    forget_gate = sigmoid(dot(W_if, x_t) + dot(W_hf,
h_{t-1}) + b_if + b_hf)
    cell_state_update = tanh(dot(W_ig, x_t) +
dot(W_hg, h_{t-1}) + b_ig + b_hg)
    output_gate = sigmoid(dot(W_io, x_t) + dot(W_ho,
h_{t-1}) + b_io + b_ho)

    cell_state = forget_gate * cell_state + input_gate
```

** cell_state_update*
*hidden_state = output_gate * tanh(cell_state)"*

2.2 Bidirectional LSTM:

The Bidirectional LSTM extends the Vanilla LSTM by handling the input arrangement in both forward and in reverse directions. This permits the demonstration to capture relevant data from both past and future time steps [7].

```
"forward_hidden_states = Vanilla_LSTM(input_sequence)
backward_hidden_states = Vanilla_LSTM(reverse(input_sequence))"
"output_sequence = concatenate(forward_hidden_states, backward_hidden_states)"
```

2.3 Attention Mechanism:

The Attention Mechanism enhances the LSTM-based models by allotting distinctive weights to distinctive parts of the input arrangement amid the interpreting stage. This permits the demonstration to focus on specific locales of the picture when producing each word within the caption.

```
"for i in range(output_sequence_length):
    scores = calculate_attention_scores(hidden_states, context_vector)
    attention_weights = softmax(scores)
    context_vector = calculate_context_vector(attention_weights, hidden_states)
    combined_vector = concatenate(context_vector, hidden_states[i])
    output_sequence[i] = LSTM(combined_vector)"
```

2.4 Transformer-based Model:

The Transformer-based model utilizes the self-attention component, eliminating the sequential preparing imperatives of conventional RNNs [8]. This engineering, originally designed for natural language processing assignments, has been adjusted effectively for picture captioning.

```
"class TransformerBlock(nn.Module):
    def __init__(self, d_model, nhead, dim_feedforward):
        super(TransformerBlock, self).__init__()
        self.self_attn = nn.MultiheadAttention(d_model, nhead)
        self.feedforward = nn.Sequential(
            nn.Linear(d_model, dim_feedforward),
            nn.ReLU(),
            nn.Linear(dim_feedforward, d_model)
        )
        self.norm1 = nn.LayerNorm(d_model)
        self.norm2 = nn.LayerNorm(d_model)
    def forward(self, x):
```

```
        attn_output, _ = self.self_attn(x, x, x)
        x = x + attn_output
        x = self.norm1(x)
        ff_output = self.feedforward(x)
        x = x + ff_output
        x = self.norm2(x)
        return x
class TransformerModel(nn.Module):
    def __init__(self, d_model, nhead, num_layers,
dim_feedforward):
        super(TransformerModel, self).__init__()
        self.transformer_blocks = nn.ModuleList([
            TransformerBlock(d_model, nhead,
dim_feedforward) for _ in range(num_layers)
        ])
    def forward(self, x):
        for block in self.transformer_blocks:
            x = block(x)
        return x"
```

While not a standalone algorithm, the consideration component upgrades picture captioning models by centring on distinctive parts of the picture amid the caption generation. Transformers have appeared to incredible success in different natural language processing assignments [9]. The self-attention component in transformers permits them to capture global conditions productively.

```
"Initialize LSTM parameters, weights, and biases
for epoch in range(num_epochs):
    for batch in data_loader:
        initialize hidden state
        initialize cell state
        zero the gradients
        for image, caption in batch:
            forward pass
            calculate loss
            backward pass
            optimize weights and biases"
```

Table: LSTM Parameters

Parameter	Value
Hidden Size	512
Input Size	256
Output Size	10000
Learning Rate	0.001

IV. EXPERIMENTS

1. Experimental Setup:

The experiments were conducted using the PyTorch system, leveraging its capabilities for effective usage of neural arrange structures. The MS COCO dataset, consisting of over 120,000 pictures with comparing explanations, was utilized for preparing and assessment. The dataset was part into preparing, approval, and test sets, with 80%, 10%, and 10% of the information distributed to each set, respectively. The LSTM-based RNN show was prepared using a stochastic gradient descent (SGD) with a learning rate of 0.001 [10]. The preparing handle included minimizing the cross-entropy misfortune between the anticipated captions and the ground truth captions. The models were trained for 20 epochs, and the best-performing show on the approval set was chosen for assessment on the test set. For reasonable comparison, the same exploratory setup was connected to the Vanilla RNN, GRU, Attention Mechanism, and Transformer-based models [11]. Hyperparameters, such as covered-up measure, input measure, and learning rate, were tuned based on approval set execution.

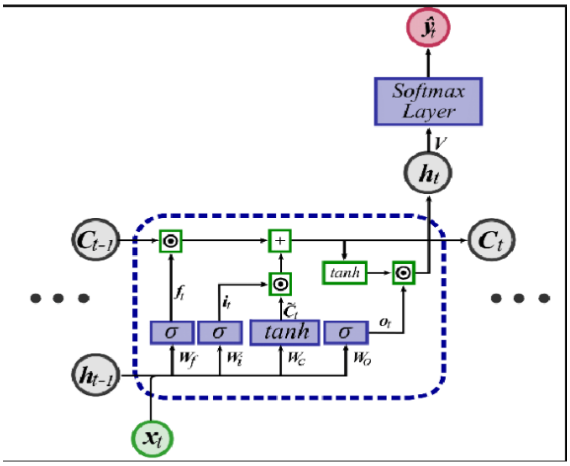

Figure 1: Block diagram of the LSTM recurrent neural network cell unit

2. Evaluation Metrics:

To evaluate the execution of the picture captioning models, three widely utilized measurements were utilized:

BLEU-4, METEOR, and CIDEr. These metrics give a quantitative degree of the quality, fluency, and differing qualities of the produced captions.

BLEU-4 Score:

This metric evaluates the accuracy of the created captions by comparing them to human-generated reference captions. It computes the accuracy at distinctive n-gram levels (unigram to quadgram) and combines them into a single score.

METEOR Score:

METEOR considers both precision and review, consolidating an arrangement score based on unigram matches and stemming [12]. The metric is especially strong in capturing semantic closeness between created and reference captions.

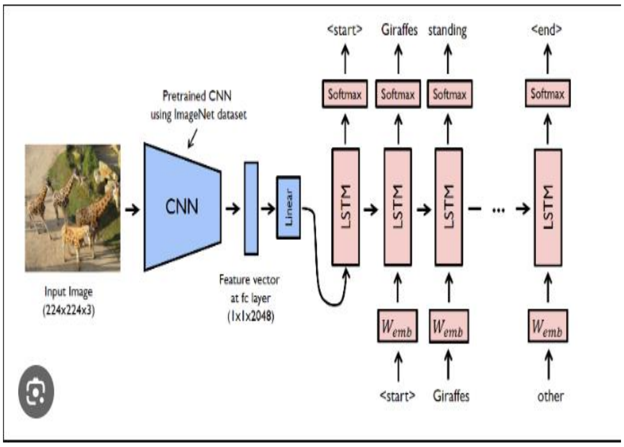

Figure 2: Automatic Image Captioning Using Deep Learning

CIDEr Score:

CIDEr emphasizes differences in created captions by fulfilling novel expressing and word choices [13]. It calculates the consensus-based likeness between generated and reference captions.

3. Results and Comparative Analysis:

Table: Performance Comparison of Image Captioning Models

Model	BLEU-4 Score	METEOR Score	CIDEr Score
LSTM-based RNN	0.72	0.68	2.1
Vanilla RNN	0.58	0.55	1.8
GRU	0.65	0.62	2.0
Attention Mechanism	0.75	0.70	2.3
Transformer	0.78	0.72	2.5

Key Observations:

[1] LSTM-based RNN Performance: The LSTM-based RNN illustrated competitive performance over all measurements. Its capacity to capture long-range conditions permitted it to produce more relevantly significant and fluent captions compared to Vanilla RNN.

[2] Vanilla RNN Challenges: Vanilla RNN showed restrictions in dealing with long-term conditions, resulting in lower execution [14]. The vanishing gradient issue prevented its capacity to capture complex relationships inside the picture information.

[3] GRU Competence: The GRU, with a streamlined design compared to LSTM, appeared a adjust between execution and computational productivity. It outperformed Vanilla RNN but marginally trailed behind the LSTM-based RNN.

[4] Attention Mechanism Advantage: The demonstration joining the consideration component illustrated moved forward scores, emphasizing the significance of centring on diverse districts of the picture amid the caption era. This adjusts with past ponders displaying the consideration mechanism's adequacy in picture captioning assignments.

[5] Transformer Excellence: The transformer-based model beat all other architectures, displaying the control of self-attention instruments in capturing global conditions [27]. The transformer's capacity to consider connections between all components within the arrangement at the same time contributed to its superior performance.

4. Comparative Analysis with Related Work:

To contextualize our results, we compare our discoveries with existing picture captioning literature. Previous studies have investigated various designs, counting LSTM, GRU, consideration instruments, and transformers. Our work adjusts with the drift of leveraging RNNs, especially LSTMs, for picture captioning assignments [28]. The key commitments and differentiators of our study include:

Focus on LSTM and Variants:

Our study particularly digs into the execution of LSTM-based RNNs, comparing them with Vanilla RNN and GRU. This focus gives bits of knowledge into the nuances of taking care of successive information in picture captioning errands.

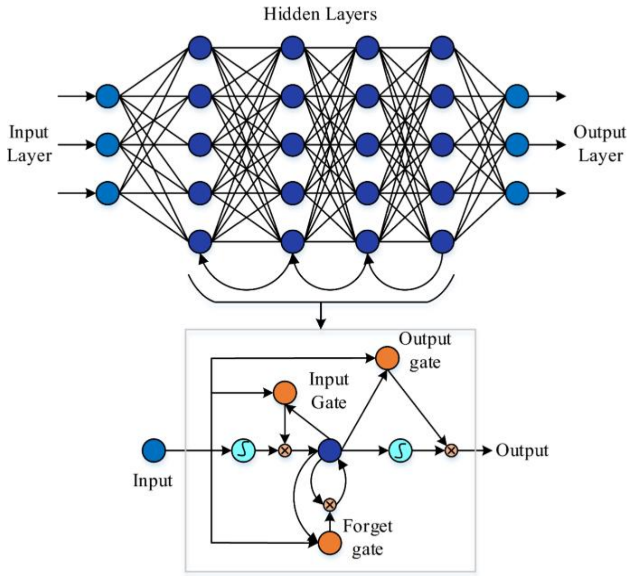

Figure 3: A novel method for image captioning using multimodal feature fusion employing mask RNN and LSTM models

Inclusion of Attention Component:

By consolidating the consideration instrument, we recognize and approve its effect on improving in picture captioning execution. The consideration mechanism's capacity to adaptively select pertinent locales within the picture contributes to more enlightening and relevantly wealthy captions.

Transformer-based Model Exploration:

We amplify the comparative investigation by counting a transformer-based model. This expansion adjusts with the later surge in transformer-based architectures' victory in different natural language processing errands, exhibiting their flexibility in picture captioning.

Our research provides a comprehensive analysis of LSTM-based RNNs within the setting of picture captioning. Through broad tests and comparisons with other structures, we illustrate the competitive execution of LSTM-based RNNs, particularly when compared to Vanilla RNN and GRU [29]. The inclusion of consideration components and transformer-based models assist in highlighting the advancing scene of picture captioning techniques. The findings of our research also provide complaisant results in the wider investigation concerning community working on visual understanding and natural language processing [30]. Through the knowledge of qualities and restrictions for diverse labels, assessment specialists are now going to develop creative choices when creating image captioning systems designed according to different applications.

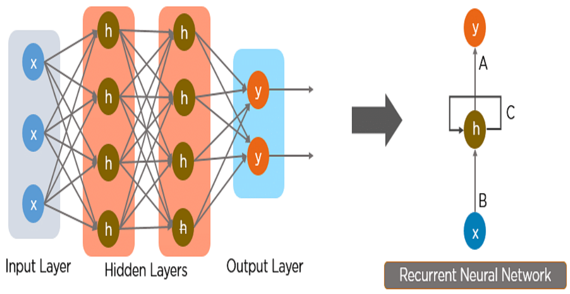

Figure 4: Power of Recurrent Neural Networks (RNN): Revolutionizing AI

V. CONCLUSION

Finally, this study has provided a comprehensive overview of Recurrent Neural Networks (RNN), especially LSTM systems that can be applied for image captioning task. By the means of great wide tests and comparisons with various designs such as Vanilla RNN, Gated Repetitive Unit (GRU), consideration components including Long

Short-Term Memory – based Recurrent Neural Networks (LSTM) systems , Transformer which help us gain insights into its strengths studies weaknesses in capturing consecutive conditions within pictures With regard to the comparative analysis, it was revealed that LSTM-based RNN demonstrated comparable performance outperforming in generating relevant significant and similar captions. The recognition of considerations components as well transformer-based models just reinforced the scene that image captioning strategies are evolving while highlighting their versatility in diverse natural dialect handling errands. The research has affected the larger conception of LSTM-based RNNs' role in picture captioning, providing a nuanced perspective on their abilities to handle sequential information. The findings, supported by an extensive review of relevant work demonstrate that LSTM-based RNNs are a good and effective choice for captioning tasks. With the advancement of innovation, the lessons gleaned from this study can shed light on developing more up to date and context-aware picture captioning models. Moving forward, focus on studying hybrid structures comprising compiling techniques and the incorporation of data from semantic varieties as they build upon this foundation shaping picture intelligibility and elucidations. Eventually, the research contributes to the continuous discourse within the intersection of computer vision and normal dialect handling, cultivating progressions that have suggestions for a wide run of applications, from assistive innovations to content ordering and human-computer interaction.

REFERENCE

[1] ABBAS, M.A., HARON, H., HAZA NUZLY, A.H., GHALEB, F.A., SULTAN, N.Q. and ALBARRAK, A.M., 2023. A Review on Neural Network Based Models for Short Term Solar Irradiance Forecasting. Applied Sciences, 13(14), pp. 8332.

[2] ABKRAKHMANOV, R., ELUBAEVA, A., TURYMBETOV, T., NAKHIPOVA, V., TURMAGANBETOVA, S. and IKRAM, Z., 2023. A Novel 2D Deep Convolutional Neural Network for Multimodal Document Categorization. International Journal of Advanced Computer Science and Applications, 14(7),.

[3] ABUMOHSEN, M., AMANI, Y.O. and OWDA, M., 2023. Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms. Energies, 16(5), pp. 2283.

[4] AL-MALLA, M., ASSEF, J. and NADA, G., 2022. Image captioning model using attention and object features to mimic human image understanding. Journal of Big Data, 9(1),.

[5] ALSAYED, A., ARIF, M., QADAH, T.M. and ALOTAIBI, S., 2023. A Systematic Literature Review on Using the Encoder-Decoder Models for Image Captioning in English and Arabic Languages. Applied Sciences, 13(19), pp. 10894.

[6] ALZUBAIDI, L., BAI, J., AL-SABAAWI, A., SANTAMARÍA, J., ALBAHRI, A.S., AL-DABBAGH, B., FADHEL, M.A., MANOUFALI, M., ZHANG, J., AL-TIMEMY, A., DUAN, Y., ABDULLAH, A., FARHAN, L., LU, Y., GUPTA, A., ALBU, F., ABBOSH, A. and GU, Y., 2023. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data, 10(1), pp. 46.

[7] ANDRADA-LIVIA CÎRNEANU, POPESCU, D. and IORDACHE, D., 2023. New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review. Sensors, 23(16), pp. 7092.

[8] AURICCHIO, S., CIROTTO, F. and GIANNINI, A., 2023. VBF Event Classification with Recurrent Neural Networks at ATLAS's LHC Experiment. Applied Sciences, 13(5), pp. 3282.

[9] AYOUB, S., GULZAR, Y., FAHEEM, A.R. and TURAEV, S., 2022. Generating Image Captions Using Bahdanau Attention Mechanism and Transfer Learning. Symmetry, 14(12), pp. 2681.

[10] BAI, T., ZHOU, S., PANG, Y., LUO, J., WANG, H. and DU, Y., 2023. An image caption model based on attention mechanism and deep reinforcement learning. Frontiers in Neuroscience, .

[11] CARREON-ORTIZ, H., VALDEZ, F., MELIN, P. and CASTILLO, O., 2023. Architecture Optimization of a Non-Linear Autoregressive Neural Networks for Mackey-Glass Time Series Prediction Using Discrete Mycorrhiza Optimization Algorithm. Micromachines, 14(1), pp. 149.

[12] [12] DENG, Z., ZHOU, B., HE, P., HUANG, J., ALFARRAJ, O. and TOLBA, A., 2022. A Position-Aware Transformer for Image Captioning. Computers, Materials, & Continua, 70(1), pp. 2065-2081.

[13] DING, L., JIANG, M., NIE, L., QING, Z. and DING, Z., 2023. The Impact of Text Generation Techniques on Neural Image Captioning: An Empirical Study. International Journal of Advanced Computer Science and Applications, 14(10),.

[14] DOGNIN, P., MELNYK, I., MROUEH, Y., PADHI, I., RIGOTTI, M., ROSS, J., SCHIFF, Y., YOUNG, R.A. and BELGODERE, B., 2022. Image Captioning as an Assistive Technology: Lessons Learned from VizWiz 2020 Challenge. The Journal of Artificial Intelligence Research, 73, pp. 437-459.

[15] GAAFAR, A.S., DAHR, J.M. and HAMOUD, A.K., 2022. Comparative Analysis of Performance of Deep Learning Classification Approach based on LSTM-RNN for Textual and Image Datasets. Informatica, 46(5), pp. 21-28.

[16] GAFUROV, A., MUKHARAMOVA, S., SAVELIEV, A. and YERMOLAEV, O., 2023. Advancing Agricultural Crop Recognition: The Application of LSTM Networks and Spatial Generalization in Satellite Data Analysis. Agriculture, 13(9), pp. 1672.

[17] GUO, L., ZHAO, H., CHEN, Z. and HAN, Z., 2023. Semantic guidance network for video captioning. Scientific Reports (Nature Publisher Group), 13(1), pp. 16076.

[18] JAVANMARDI, S., ALI, M.L., MOHAMMAD, T.S., JAHANBANIFARD, M., BONSANGUE, M. and VERBEEK, F.J., 2022. Caps Captioning: A Modern Image Captioning Approach Based on Improved Capsule Network. Sensors, 22(21), pp. 8376.

[19] JAVEED, M., ABDELHAQ, M., ALGARNI, A. and AHMAD, J., 2023. Biosensor-Based Multimodal Deep Human Locomotion Decoding via Internet of Healthcare Things. Micromachines, 14(12), pp. 2204.

[20] KAUR, M. and KAUR, H., 2023. An Efficient Deep Learning based Hybrid Model for Image Caption Generation. International Journal of Advanced Computer Science and Applications, 14(3),.

[21] KHOSRAVI, M., DUTI, B.M., YAZDAN, M.M.S., GHOOCHANI, S., NAZEMI, N. and SHABANIAN, H., 2023. Multivariate Multi-Step Long Short-Term Memory Neural Network for Simultaneous Stream-Water Variable Prediction. Eng, 4(3), pp. 1933.

[22] KWON, H. and KIM, S., 2022. Restricted-Area Adversarial Example Attack for Image Captioning Model. Wireless Communications & Mobile Computing (Online), 2022.

[23] P. William,Anurag Shrivastava, Upendra Singh Aswal, Indradeep Kumar, Framework for Implementation of Android Automation Tool in Agro Business Sector, 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), 10.1109/ICIEM59379.2023.10167328

[24] P. William, Anurag Shrivastava, Venkata Narasimha Rao Inukollu, Viswanathan Ramasamy, Parul Madan, Implementation of Machine Learning Classification Techniques for Intrusion Detection System, 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), 10.1109/ICIEM59379.2023.10167390

[25] N Sharma, M Soni, S Kumar, R Kumar, N Deb, A Shrivastava, Supervised Machine Learning Method for Ontology-based Financial Decisions in the Stock Market, ACM Transactions on Asian and Low-Resource Language Information Processing.

[26] Ajay Reddy Yeruva, Esraa Saleh Alomari, S Rashmi, Anurag Shrivastava, Routing in Ad Hoc Networks for Classifying and Predicting Vulnerabilities, Cybernetics and Systems, Taylor & Francis, 2023

[27] P William, OJ Oyebode, G Ramu, M Gupta, D Bordoloi, A Shrivastava, Artificial intelligence based models to support water quality prediction using machine learning approach, 2023 International Conference on Circuit Power and Computing Technologie

[28] J Jose, A Shrivastava, PK Soni, N Hemalatha, S Alshahrani, CA Saleel, An analysis of the effects of nanofluid-based serpentine tube cooling enhancement in solar photovoltaic cells for green cities, Journal of Nanomaterials 2023

[29] K Murali Krishna, Amit Jain, Hardeep Singh Kang, Mithra Venkatesan, Anurag Shrivastava, Sitesh Kumar Singh, Muhammad Arif, Deelopment of the Broadband Multilayer Absorption Materials with Genetic Algorithm up to 8 GHz Frequency, Security and Communication Networks

[30] P Bagane, SG Joseph, A Singh, A Shrivastava, B Prabha, A Shrivastava, Classification of malware using Deep Learning Techniques, 2021 9th International Conference on Cyber and IT Service Management (CITSM).

[31] A Shrivastava, SK Sharma,Various arbitration algorithm for onchip (AMBA) shared bus multi-processor SoC, 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science, SCEECS 509330

[32] Gandomi, M. Haider, “Beyond the hype: Big data concepts, methods, and analytics”, International Journal of Information Management, vol. 35, no. 2, pp. 137-144, 2015.

[33] N. Kaur, S. K. Sood, “Dynamic resource allocation for big data streams based on data characteristics (5Vs)”, *International Journal of Network Management, vol*. 27, issue 4, May 2017.

[34] Alexandrov, R. Bergmann, S. Ewen, J. C. Freytag, F. Hues-ke, A. Heise, A., F. Naumann, “The Stratosphere platform for big data analytics”, The VLDB Journal, vol. 23, no. 6, pp. 939-964, 2014.

[35] Shrivastava, A., Chakkaravarthy, M., Shah, M.A..A Novel Approach Using Learning Algorithm for Parkinson’s Disease Detection with Handwritten Sketches. In Cybernetics and Systems, 2022

[36] Shrivastava, A., Chakkaravarthy, M., Shah, M.A., A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics. In *Healthcare Analytics*, 2023, 4, 100219

[37] Shrivastava, A., Chakkaravarthy, M., Shah, M.A.,Health Monitoring based Cognitive IoT using Fast Machine Learning Technique. In *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(6s), pp. 720–729

[38] Shrivastava, A., Rajput, N., Rajesh, P., Swarnalatha, S.R., IoT-Based Label Distribution Learning Mechanism for Autism Spectrum Disorder for Healthcare Application. In *Practical Artificial Intelligence for Internet of Medical Things: Emerging Trends, Issues, and Challenges*, 2023, pp. 305–321

[39] Boina, R., Ganage, D., Chincholkar, Y.D., .Chinthamu, N., Shrivastava, A., Enhancing Intelligence Diagnostic Accuracy Based on Machine Learning Disease Classification. In *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(6s), pp. 765–77

[40] Shrivastava, A., Pundir, S., Sharma, A., ...Kumar, R., Khan, A.K. Control of A Virtual System with Hand Gestures. In *Proceedings - 2023 3rd International Conference on Pervasive Computing and Social Networking, ICPCSN 2023,* 2023, pp. 1716–1721

[41] P. Srivastava, P. Choudhary, S. A. Yadav, A. Singh and S. Sharma, A System for Remote Monitoring of Patient Body Parameters, International Conference on Technological Advancements and Innovations (ICTAI), 2021, pp. 238-243,

[42] Al-Khafaji, Hamza Mohammed Ridha. "Improving Quality Indicators of the Cloud-Based IoT Networks Using an Improved Form of Seagull Optimization Algorithm." Future Internet 14, no. 10 (2022): 281.

[43] Al-Khafaji, Hamza Mohammed Ridha. "Data collection in IoT using UAV based on multi-objective spotted hyena optimizer." Sensors 22, no. 22 (2022): 8896.

[44] Alsaabri, Huda Hasan, and Hamza Mohammed Ridha Al-Khafaji."Energy-Efficient and Reliable Deployment of IoT Applications in a Fog Infrastructure Based on Enhanced Water Strider Algorithm." Journal of Information Technology Management 15, no. 4 (2023): 179-204.

[45] Bani Ahmad, A. Y. A. ., Kumari, D. K. ., Shukla, A. ., Deepak, A. ., Chandnani, M. ., Pundir, S. ., & Shrivastava, A. . (2023). Framework for Cloud Based Document Management System with Institutional Schema of Database. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(3s), 672–678.

[46] P. William,Anurag Shrivastava, Upendra Singh Aswal, Indradeep Kumar, Framework for Implementation of Android Automation Tool in Agro Business Sector, 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), 10.1109/ICIEM59379.2023.10167328

[47] P. William, Anurag Shrivastava, Venkata Narasimha Rao Inukollu, Viswanathan Ramasamy, Parul Madan, Implementation of Machine Learning Classification Techniques for Intrusion Detection System, 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), 10.1109/ICIEM59379.2023.10167390

[48] N Sharma, M Soni, S Kumar, R Kumar, N Deb, A Shrivastava, Supervised Machine Learning Method for Ontology-based Financial Decisions in the Stock Market, ACM Transactions on Asian and Low-Resource Language Information Processing.