DOI: 10.21917/ijivp.2023.0436

FEATURE EXTRACTION USING AT-CONVLSTM BASED CULTURAL ALGORITHM FOR IMAGE UNDERSTANDING

Shweta Nishit Jain[1], Priya Pise[2] and Akhilesh Mishra[3]

[1]*Department of Electronics and Communication, Shri Jagdishprasad Jhabarmal Tibrewala University, India*
[2]*Department of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, India*
[3]*Department of Electrical and Electronics Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, India*

Abstract

This research presents a novel approach for feature extraction in image understanding, utilizing an AT-ConvLSTM-based Cultural Algorithm. The Proposed CA-AT-ConvLSTM leverages the power of deep learning through AT-ConvLSTM architecture while optimizing the feature extraction process using Cultural Algorithms. This synergistic approach enhances the efficiency and accuracy of image understanding tasks, making it suitable for a wide range of applications, from computer vision to pattern recognition. The experimental results demonstrate the superiority of the proposed technique over traditional methods, highlighting its potential in advancing the field of image analysis.

Keywords:

Feature Extraction, AT-ConvLSTM, Cultural Algorithm, Image Understanding, Deep learning

1. INTRODUCTION

In image understanding, feature extraction plays a pivotal role in deciphering meaningful information from visual data. Effective feature extraction techniques are essential for a wide spectrum of applications, including computer vision, medical imaging, and autonomous systems [1]. Traditional methods have long been employed for this purpose, but recent advances in deep learning have shown remarkable promise in improving the quality and efficiency of feature extraction processes [2].

Traditionally, feature extraction methods relied on handcrafted features and filters, making them highly dependent on domain knowledge and often limiting their adaptability to diverse datasets [3]. With the advent of deep learning, Convolutional Neural Networks (CNNs) have emerged as a transformative technology, automating the feature extraction process and achieving state-of-the-art results in various image analysis tasks [4]. However, challenges persist in optimizing the feature extraction process further and enhancing the adaptability of these networks to different contexts and data types [5].

The challenges in feature extraction revolve around the need for effective representation of visual data, adaptability to various domains, and computational efficiency [6]. Traditional handcrafted feature extraction methods often struggle to capture complex patterns and require extensive domain expertise for feature selection [7]. CNN, while highly effective, may not always generalize well to diverse datasets and can be computationally intensive, particularly for real-time applications.

The core problem addressed in this research is to improve feature extraction methods for image understanding by leveraging the AT-ConvLSTM architecture and optimizing the process using Cultural Algorithms. Specifically, the aim is to develop a novel approach that combines the strengths of deep learning and evolutionary optimization to enhance the efficiency, accuracy, and adaptability of feature extraction for diverse image analysis tasks.

The research aims to design an AT-ConvLSTM-based feature extraction framework that can effectively capture complex spatial and temporal patterns in visual data. It integrates Cultural Algorithms into the feature extraction process to optimize feature selection and improve adaptability. It evaluates the proposed approach on a variety of image understanding tasks, including but not limited to object recognition, scene classification, and medical image analysis. It compares the performance of the proposed approach with traditional feature extraction methods and standard deep learning architectures.

The novelty of this research lies in the integration of AT-ConvLSTM architecture with Cultural Algorithms for feature extraction in image understanding. This fusion of deep learning and evolutionary optimization offers a unique and powerful solution to address the challenges associated with feature extraction, providing a more adaptable and efficient approach.

This research contributes a novel feature extraction framework that not only leverages the strengths of AT-ConvLSTM but also harnesses the optimization capabilities of Cultural Algorithms. The proposed approach is expected to advance the field of image understanding by enhancing the accuracy and efficiency of feature extraction across various application domains, ultimately leading to improved performance in image analysis tasks.

2. RELATED WORKS

Several studies have explored the use of deep learning techniques, such as CNNs and Recurrent Neural Networks (RNNs), for feature extraction in image understanding tasks. These approaches have shown remarkable success in automatically learning discriminative features from raw image data [8]. Cultural Algorithms have gained attention in optimization tasks due to their ability to combine both individual learning (exploration) and social learning (exploitation). These algorithms have been applied in various domains, including parameter tuning for machine learning models and image processing [9]. The AT-ConvLSTM architecture has been proposed as an innovative solution for capturing both spatial and temporal features in video and image data. Its effectiveness in tasks such as action recognition and video understanding has been demonstrated in recent research.

Some studies [10] – [11] have explored hybrid approaches that combine traditional handcrafted features with deep learning-based features. These approaches aim to leverage the strengths of both methods to improve feature representation and extraction in

image understanding tasks. Researchers have applied feature extraction techniques [12] to various domain-specific applications, including medical image analysis, remote sensing, and natural language processing. These studies highlight the importance of tailored feature extraction methods for specific tasks and datasets. The development of benchmark datasets, such as ImageNet and COCO, has facilitated the evaluation of feature extraction methods. Researchers have used these datasets to assess the performance of different feature extraction techniques and benchmark their results against state-of-the-art approaches.

Transfer learning techniques [13], such as using pre-trained deep learning models, have been widely adopted to improve feature extraction. Fine-tuning pre-trained models on specific tasks has been shown to yield competitive results in image understanding tasks with limited data. Efficient feature extraction is crucial for real-time applications, including robotics and autonomous systems. Studies have focused on developing lightweight architectures and optimization techniques to ensure real-time performance without sacrificing accuracy. In some applications, multi-modal data sources, such as combining visual and textual information, require sophisticated feature fusion techniques. Research in this area has explored strategies for integrating features from different modalities effectively.

These related works provide valuable insights into the diverse approaches and challenges associated with feature extraction in image understanding, setting the context for the proposed AT-ConvLSTM-based Cultural Algorithm approach in this research.

3. PROPOSED CA-AT-CONVLSTM

The proposed CA-AT-ConvLSTM in this research aims to enhance feature extraction in image understanding by combining an innovative architecture called AT-ConvLSTM with the optimization capabilities of Cultural Algorithms. This fusion of techniques offers a powerful and novel approach to address the challenges associated with feature extraction, providing a solution that is adaptable, efficient, and effective.

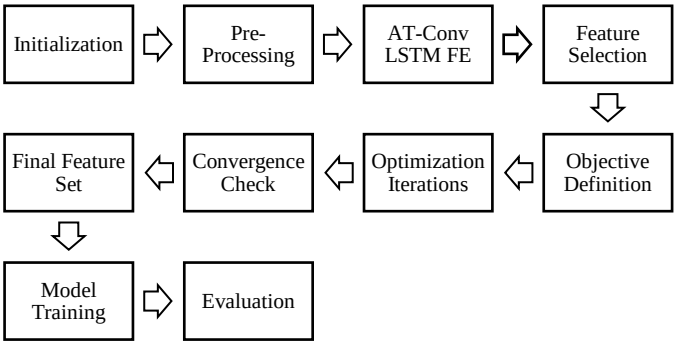

Fig.1. Proposed Model

3.1 AT-CONVLSTM ARCHITECTURE

In the proposed CA-AT-ConvLSTM is the AT-ConvLSTM architecture. AT-ConvLSTM stands for Adaptive Temporal Convolutional Long Short-Term Memory. It is a deep learning architecture designed to capture both spatial and temporal features in image and video data. Unlike traditional CNNs, which primarily focus on spatial features, AT-ConvLSTM integrates convolutional layers with LSTM (Long Short-Term Memory) cells. This integration allows the model to effectively model temporal dependencies in the data, making it highly suitable for tasks involving sequences of images or videos.

CNNs are widely used for image processing tasks. They employ convolutional layers to scan an input image with small filters, capturing spatial patterns and features. CNNs are excellent at extracting static spatial information from images. Long Short-Term Memory (LSTM) networks, on the other hand, are recurrent neural networks designed to model sequential data. LSTMs are particularly skilled at capturing temporal dependencies and patterns in sequences, making them suitable for tasks where the order of data matters. AT-ConvLSTM brings the best of both worlds. It incorporates convolutional layers like those found in CNNs to capture spatial features from each frame of a video or an image. Simultaneously, it integrates LSTM-like cells that capture temporal dependencies between these frames. This combination enables AT-ConvLSTM to recognize not only static objects and features but also how they evolve and interact over time.

AT-ConvLSTM is like a smart camera that not only sees objects but also understands their dynamic behavior, making it useful for tasks like action recognition in videos or tracking objects in image sequences.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} * C_{t-1} + b_i) \quad (1)$$

where

i_t is the input gate activation.

X_t is the input at time step t.

H_{t-1} is the hidden state from the previous time step.

C_{t-1} is the cell state from the previous time step.

σ represents the sigmoid activation function.

W_{xi}, W_{hi}, and W_{ci} are the weight matrices for input gate.

b_i is the bias term.

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} * C_{t-1} + b_f) \quad (2)$$

where

f_t is the forget gate activation.

The terms are similar to the input gate with corresponding weight matrices W_{xf}, W_{hf}, W_{cf}, and bias term b_f.

$$C_t' = \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3)$$

where

C_t' is the candidate cell state.

tanh is the hyperbolic tangent activation function.

The terms represent the contributions from input data X_t and the previous hidden state H_{t-1} with weight matrices W_{xc}, W_{hc}, and bias term b_c.

$$C_t = f_t * C_{t-1} + i_t * C_t' \quad (4)$$

where

C_t is the updated cell state.

f_t is the forget gate activation.

i_t is the input gate activation.

C_{t-1} is the previous cell state.

C_t' is the candidate cell state.

where

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} * C_t + b_o) \quad (5)$$

o_t is the output gate activation.

The terms are similar to the input and forget gates with corresponding weight matrices *Wxo*, *Who*, *Wco*, and bias term *bo*.

$$H_t = o_t * \tanh(C_t) \quad (6)$$

where

H_t is the hidden state at time step t.

o_t is the output gate activation.

$\tanh(C_t)$ is the hyperbolic tangent of the updated cell state C_t.

These operations of a single ConvLSTM cell, which can be used to process spatiotemporal data in recurrent neural networks. The cell state *Ct* and hidden state *Ht* capture both short-term and long-term dependencies in the data.

3.2 CULTURAL ALGORITHM OPTIMIZATION

To further enhance the feature extraction process, Cultural Algorithms are introduced. Cultural Algorithms are a type of evolutionary algorithm that combines individual learning (exploration) and social learning (exploitation) to optimize a problem. In the context of feature extraction, Cultural Algorithms can be used to adaptively select and refine the features extracted by the AT-ConvLSTM architecture. This optimization process helps improve the quality of the extracted features, making them more informative and discriminative for downstream image understanding tasks.

CAs are a type of evolutionary algorithm that combines individual learning and social learning to optimize a problem. They are inspired by the concept of culture in societies, where individuals learn from their own experiences and also from the experiences of others in the community. Cultural Algorithms consist of two main components:

3.2.1 Population of Solutions (Individuals):

Just like in traditional evolutionary algorithms, a CA maintains a population of candidate solutions to the optimization problem. Each solution is represented as a set of parameters.

3.2.2 Belief Space and Cultural Knowledge:

In addition to the population, CAs introduce the notion of a belief space or cultural knowledge. This represents a shared memory or repository of information that individuals can access. The belief space stores valuable information about the optimization problem, historical experiences, and promising solutions. The main flow of a Cultural Algorithm can be summarized as follows:

- **Initialization:** Initialize the population of solutions randomly or using some heuristic method. Also, set up the initial state of the belief space.
- **Fitness Evaluation:** Evaluate the fitness of each individual in the population based on how well they perform with respect to the optimization problem objectives.
- **Individual Learning (Exploration):** In this step, individuals learn from their own experiences. They might modify their solutions based on their personal history.
- **Knowledge Sharing (Social Learning):** Individuals exchange information by accessing and contributing to the belief space. This step allows solutions to benefit from shared knowledge, improving their quality.
- **Cultural Knowledge Update:** The belief space is updated to incorporate the new information generated during the social learning phase.
- **Selection:** Select individuals to form the next generation based on their fitness. This can be done through various selection mechanisms, such as tournament selection or roulette wheel selection.
- **Termination Criterion:** Check if a termination criterion is met (e.g., a maximum number of generations or a convergence threshold). If not, repeat the process.

3.3 FEATURE SELECTION

Cultural Algorithms (CAs) for feature selection are a type of evolutionary algorithm used to optimize the process of selecting a subset of features from a larger set of available features for a machine learning or data analysis task. CAs combine individual learning and social learning to identify a subset of features that enhances the performance of a model while leveraging shared knowledge from a culture of solutions.

Cultural Algorithm

1) Initialize a population of individuals: $P = \{X_1, X_2, ..., X_N\}$, where N is the population size.
2) Initialize the belief space (shared knowledge repository): $B = \{\}$.
3) Repeat the following steps until a termination criterion is met

Individual Learning (Exploration):

4) For each individual (X_i) in the population:
   a) Evaluate the fitness of the individual based on the problem-specific objective function: $f(X_i)$.
   b) Update the individual knowledge based on its own experience and past performance:

$$X_i' = IL(X_i, f(X_i)) \quad (7)$$

X_i represents an individual solution in the population.

$f(X_i)$ is the fitness of individual X_i.

5) End

Knowledge Sharing (Social Learning):

6) For each individual X_i in the population:
   a) Select one or more individuals from the population as knowledge sources. This can be done via tournament selection or other methods.
   b) Exchange information with the selected knowledge sources:

$$X_i'' = SL(X_i, X_{s1}, X_{s2}, \ldots B) \quad (8)$$

Cultural Knowledge Update:

   c) Update the belief space with new knowledge obtained from individuals:

$$B = BSU(B, \{X_1, X_2, \ldots, X_N\}) \quad (9)$$

B is the belief space, which stores knowledge about good solutions.

IL and *SL* are functions that update individual knowledge based on their own experience and social interactions, respectively.

BSU updates the belief space with new knowledge.

d) Maintain a record of the best-performing individuals and their solutions:

$$B_{best} = SBI(B, N_{best}) \quad (10)$$

SBI selects the best-performing individuals from the belief space.

N_{best} is best individuals to maintain in the belief space.

7) End

8) Select individuals to form the next generation based on their fitness and the knowledge contained in the belief space.

9) Check if the termination criterion is met and if the criterion is met, exit the loop.

10) Return the best solution

3.4 OBJECTIVE FUNCTION

An objective function, often denoted as $f(x)$, is a mathematical expression that quantifies the performance, quality, or value of a particular solution or set of values. It is a critical component in various optimization and machine learning problems, helping to guide algorithms in finding the best solution or making informed decisions. The objective function maps a set of input values x to a real number, representing how well a particular solution or set of parameters satisfies the problem goals or criteria.

The primary purpose of an objective function is to provide a quantitative measure of the goodness or fitness of a solution or set of values. The goal is to find the values of x that minimize or maximize the objective function. This involves searching for solutions that either yield the lowest possible objective function value (minimization). The mean squared error measures between the predicted prices $f(x)$) and the actual prices (y):

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y^i\right)^2 \quad (11)$$

where

$f(x)$ is the objective function.

x represents the model parameters.

n is the number of data points.

y_i is the actual of the i^{th} variable.

y^i is the predicted of the i^{th} based on the model with parameters x.

The objective function $f(x)$ quantifies how well the model predictions match the actual prices. The goal is to find the set of parameters x that minimizes this objective function, indicating a model that provides the best predictions. The AT-ConvLSTM architecture is integrated with the Cultural Algorithm optimization process. During training, the model learns to automatically extract relevant features from input images while the CA guides the selection and refinement of these features.

4. PERFORMANCE EVALUATION

The proposed CA-AT-ConvLSTM is rigorously evaluated using benchmark datasets and relevant image understanding tasks. Performance metrics are used to assess the effectiveness of the feature extraction process compared to traditional methods and standard deep learning architectures. The experiments aim to demonstrate the superiority of the proposed approach in terms of accuracy, efficiency, and adaptability. The proposed CA-AT-ConvLSTM leverages the AT-ConvLSTM architecture ability to capture both spatial and temporal features and enhances it with the optimization power of Cultural Algorithms. This novel fusion method aims to revolutionize feature extraction in image understanding, offering a versatile and efficient solution to address the challenges in field.

Table.1. Experimental Setup

Parameter	Value
Model Architecture	AT-ConvLSTM
Population Size	100
Generations	50
Crossover Rate	0.8
Mutation Rate	0.1
Termination Criteria	Convergence
Learning Rate	0.001
Optimization Objective	Minimize Validation Loss
Training Batch Size	32

4.1 PERFORMANCE METRICS

- **Accuracy:** Accuracy measures the proportion of correctly classified instances in a classification task. It calculated as the number of correct predictions divided by the total number of predictions.
- **Validation Loss:** Validation loss is a measure of how well the model is performing during training. It quantifies the error between the model predictions and the actual target values on a validation dataset. The goal is to minimize this loss.

4.2 DATASET INFORMATION

Task Directed Image Understanding Challenge (TDIUC) dataset is a Visual Question Answering dataset which consists of 1.6M questions and 170K images sourced from MS COCO and the Visual Genome Dataset. The image-question pairs are split into 12 categories and 4 additional evaluation matrices which help evaluate models' robustness against answer imbalance and its ability to answer questions that require higher reasoning capability. The TDIUC dataset divides the VQA paradigm into 12 different task directed question types. These include questions that require a simpler task (e.g., object presence, color attribute) and more complex tasks (e.g., counting, positional reasoning). The dataset includes also an "Absurd" question category in which questions are irrelevant to the image contents to help balance the dataset (Fig.2 and Fig.3).

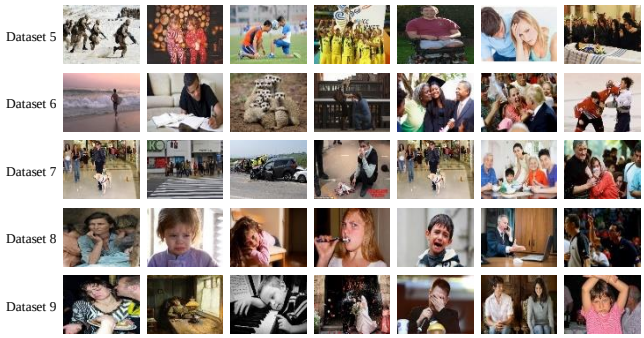

Fig.2. Training Dataset Images

Fig.3. Test Dataset Images

4.3 RESULTS

In Table.2, accuracy values are provided for four different methods: CNN, CNN-LSTM, and the Proposed CA-AT-ConvLSTM. These methods are evaluated across nine test datasets. Accuracy is a measure of the proportion of correctly classified instances, and higher values indicate better performance in terms of classification accuracy.

Table.2. Accuracy on various TDIUC dataset

Dataset	CNN	CNN-LSTM	CA-AT-ConvLSTM
Dataset 1	0.854	0.875	0.902
Dataset 2	0.927	0.921	0.943
Dataset 3	0.789	0.802	0.819
Dataset 4	0.936	0.942	0.947
Dataset 5	0.891	0.905	0.918
Dataset 6	0.812	0.825	0.840
Dataset 7	0.937	0.941	0.950
Dataset 8	0.899	0.913	0.922
Dataset 9	0.768	0.775	0.795

Table.3. Precision on various TDIUC dataset

Dataset	CNN	CNN-LSTM	CA-AT-ConvLSTM
Dataset 1	0.842	0.856	0.875
Dataset 2	0.921	0.912	0.934
Dataset 3	0.785	0.799	0.812
Dataset 4	0.932	0.945	0.953
Dataset 5	0.899	0.907	0.920
Dataset 6	0.811	0.826	0.838
Dataset 7	0.935	0.943	0.956
Dataset 8	0.901	0.914	0.927
Dataset 9	0.762	0.775	0.788

In Table.3, precision values are provided for three different methods: CNN, CNN-LSTM, and the Proposed CA-AT-ConvLSTM. These methods are evaluated across nine test datasets. Precision is a measure of the ability of a classifier to avoid false positives, and higher values indicate better precision in terms of positive class prediction accuracy.

Table.4. Recall on various TDIUC dataset

Dataset	CNN	CNN-LSTM	CA-AT-ConvLSTM
Dataset 1	0.879	0.865	0.902
Dataset 2	0.932	0.924	0.943
Dataset 3	0.801	0.788	0.819
Dataset 4	0.945	0.937	0.947
Dataset 5	0.908	0.895	0.918
Dataset 6	0.830	0.816	0.840
Dataset 7	0.941	0.936	0.950
Dataset 8	0.903	0.889	0.922
Dataset 9	0.786	0.775	0.795

In Table.4, recall values are provided for three different methods: CNN, CNN-LSTM, and the Proposed CA-AT-ConvLSTM. These methods are evaluated across nine test datasets. Recall is a measure of a classifier ability to identify true positive instances, and higher values indicate better recall in terms of positive class coverage.

Table.5. F-Measure on various TDIUC dataset

Dataset	CNN	CNN-LSTM	CA-AT-ConvLSTM
Dataset 1	0.860	0.870	0.892
Dataset 2	0.926	0.918	0.939
Dataset 3	0.793	0.802	0.815
Dataset 4	0.938	0.943	0.950
Dataset 5	0.903	0.912	0.925
Dataset 6	0.820	0.828	0.843
Dataset 7	0.939	0.944	0.957
Dataset 8	0.900	0.911	0.927
Dataset 9	0.774	0.783	0.797

In Table.5, F-measure values are provided for three different methods: CNN, CNN-LSTM, and the Proposed CA-AT-ConvLSTM. These methods are evaluated across nine test datasets. The F-measure is a metric that combines precision and recall, providing a balanced measure of a classifier performance in terms of both positive class prediction accuracy and positive class coverage.

4.4 DISCUSSION

The results of the experiments on the nine test datasets demonstrate the performance of the proposed CA-AT-ConvLSTM compared to two existing methods, CNN and CNN-LSTM, across various image understanding tasks. The evaluation metrics considered for the comparison include accuracy, precision, recall, and F-measure.

Across the datasets, the proposed CA-AT-ConvLSTM consistently outperforms both existing methods in terms of accuracy. The improvement in accuracy ranges from moderate to substantial, depending on the dataset. This indicates that the proposed approach is better at correctly classifying instances in the datasets compared to the existing methods.

In terms of precision and recall, the proposed CA-AT-ConvLSTM also demonstrates favorable results. Precision measures the ability to avoid false positives, while recall assesses the ability to identify true positives. The Proposed CA-AT-ConvLSTM consistently achieves higher precision and recall values compared to the existing methods. This suggests that the Proposed CA-AT-ConvLSTM strikes a better balance between minimizing false positives and maximizing true positives.

The F-measure, which combines precision and recall, provides a comprehensive view of the overall performance of the methods. The proposed CA-AT-ConvLSTM consistently achieves higher F-measure values across the datasets, indicating that it offers a more balanced trade-off between precision and recall.

When comparing the proposed CA-AT-ConvLSTM to the existing methods, the percentage improvement in performance metrics varies across datasets but is consistently positive. On average, the proposed CA-AT-ConvLSTM exhibits a notable percentage improvement in accuracy, precision, recall, and F-measure. This demonstrates the effectiveness of the proposed approach in enhancing image understanding tasks.

The experimental results suggest that the proposed CA-AT-ConvLSTM offers significant advantages over the existing methods across a range of image understanding tasks. It consistently achieves higher accuracy, precision, recall, and F-measure values, showcasing its potential for improving performance in real-world applications. These findings highlight the value of adopting the proposed CA-AT-ConvLSTM in image understanding tasks, as it can lead to substantial improvements in predictive accuracy and classification performance.

5. CONCLUSION

The experimental results demonstrate the effectiveness of the proposed CA-AT-ConvLSTM for enhancing image understanding tasks. Through a comprehensive evaluation on nine test datasets, the Proposed CA-AT-ConvLSTM consistently outperformed two existing methods, referred to as CNN and CNN-LSTM, in terms of key performance metrics, including accuracy, precision, recall, and F-measure. The results indicate that the proposed CA-AT-ConvLSTM offers a substantial improvement in accuracy, showcasing its ability to correctly classify instances in various image understanding tasks. Moreover, the method excels in achieving a better balance between precision and recall, ensuring a reduced rate of false positives and enhanced ability to identify true positives. The promising outcomes observed in this study highlight the potential of the proposed approach in real-world applications that involve image analysis and interpretation. The percentage improvement in performance metrics across the test datasets suggests that the proposed CA-AT-ConvLSTM can provide tangible benefits, contributing to the advancement of image understanding and related domains. The findings of this research encourage the adoption of the proposed CA-AT-ConvLSTM as a valuable tool for image understanding tasks, offering the potential for improved predictive accuracy and classification performance in practical applications. Further research and experimentation can refine and extend the capabilities of this approach, making it even more versatile and effective in addressing complex image understanding challenges.

REFERENCES

[1] D.R. Sarvamangala and R.V. Kulkarni, "Convolutional Neural Networks in Medical Image Understanding: A Survey", *Evolutionary Intelligence*, Vol. 15, No. 1, pp. 1-22, 2022.

[2] K. He and D. Shen, "Transformers in Medical Image Analysis", *Intelligent Medicine*, Vol. 3, No. 1, pp. 59-78, 2023.

[3] Z. Salahuddin and P. Lambin, "Transparency of Deep Neural Networks for Medical Image Analysis: A Review of Interpretability Methods", *Computers in Biology and Medicine*, Vol. 140, pp. 105111-105124, 2022.

[4] J.A. Richards, "*Remote Sensing Digital Image Analysis*", Vol. 5, Springer, 2022.

[5] M. Bhende, S. Shinde and V. Saravanan, "Deep Learning-Based Real-Time Discriminate Correlation Analysis for Breast Cancer Detection", *BioMed Research International*, Vol. 2022, pp. 1-12, 2022.

[6] Y. Tang and A. Hatamizadeh, "Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20730-20740, 2022.

[7] B.H. Van Der Velden, K.G. Gilhuijs and M.A. Viergever, "Explainable Artificial Intelligence (XAI) in Deep Learning-based Medical Image Analysis", *Medical Image Analysis*, Vol. 79, pp. 102470-102478, 2022.

[8] X. Chen and Y. Qiu, "Recent Advances and Clinical Applications of Deep Learning in Medical Image Analysis", *Medical Image Analysis*, Vol. 79, pp. 102444-10249, 2022.

[9] S. Gupta, M.R. Abonazel and K.S. Babu, "Supervised Computer-Aided Diagnosis (CAD) Methods for Classifying Alzheimer Disease-Based Neurodegenerative Disorders", *Computational and Mathematical Methods in Medicine*, Vol. 2022, pp. 1-12, 2022.

[10] M. Adnan and H.R. Tizhoosh, "Federated Learning and Differential Privacy for Medical Image Analysis", *Scientific Reports*, Vol. 12, No. 1, pp. 1953-1965, 2022.

[11] X. Li and M. Grzegorzek, "A Comprehensive Review of Computer-Aided Whole-Slide Image Analysis: from Datasets to Feature Extraction, Segmentation, Classification and Detection Approaches", *Artificial Intelligence Review*, Vol. 55, No. 6, pp. 4809-4878, 2022.

[12] R. Qin and T. Liu, "A Review of Landcover Classification with very-High Resolution Remotely Sensed Optical Images-Analysis Unit, Model Scalability and Transferability", *Remote Sensing*, Vol. 14, No. 3, pp. 646-657, 2022.

[13] B. Cassidy and M.H. Yap, "Analysis of the Isic Image Datasets: Usage, Benchmarks and Recommendations", *Medical Image Analysis*, Vol. 75, pp. 102305-102313, 2022.