Wheat Diseases Recognition Using Optimal Features Assisted Modified Soft Attention Network

1st Muhammad Nadeem
Department of Computer Science and Engineering
Sejong Uviversity
Seoul, Republic of Korea
muhammad.nadeem23064@gmail.com

2nd Aqib Khan
Department of Botany
Islamia College University
Peshawar 25000, Pakistan
aqibianz@gmail.com

3rd Ji-Won Kim
Department of Artificial Intelligence
Sejong University
Seoul, Republic of Korea
kimjiwonl5124@gmail.com

4th L. Minh Dang
Department of Artificial Intelligence
Sejong University
Seoul, Republic of Korea
minhdl@sejong.ac.kr

5th Hyeonjoon Moon*
Department of Computer Science and Engineering
Sejong Uviversity
Seoul, Republic of Korea
hmoon@sejong.ac.kr

***Abstract*—Wheat ranks as the third most extensively cultivated and consumed crop globally. However, many wheat crops are susceptible to spoilage caused by diverse disease types. Thus, the manual diagnosis of these diseases poses significant challenges for farmers and policymakers. Automated wheat disease recognition can potentially enhance crop yield quantity and quality. The current attempts proposed deep learning methods to predict diseases. However, they are unable to find the optimal features and attention models to focus on the diseased regions in challenging data. To overcome the research gap in the previous studies in this paper, we collected more challenging wheat disease samples to ensure the model's generalization. Then, we comparatively studied the CNN-based models to find the optimal features and specific higher representative layers. To assist the backbone features, soft attention modules are progressively modified to focus on the critical regions. The modified soft attention mechanism assists the network in prioritizing and highlighting essential regions in the wheat disease images. The findings revealed that the network with MSA outperformed the baseline model. Extensive experiments are conducted to evaluate the proposed network's superiority on different data splits.**

***Index Terms*—Wheat disease, InceptionResNetV2, Soft-attention, disease identification, wheat classification**

I. Introduction

Wheat is the most globally consumed food crop, fulfilling a significant portion of daily human energy requirements [1]. Its genetic components, particularly proteins, fiber, and vitamins, contribute to its widespread consumption [2]–[6]. Despite producing over 700 million metric tons of wheat globally in the last nine years, the demand continues to outstrip supply. It is observable that challenges like natural disasters, climate change, conflicts, and crop diseases severely impact production. Among these challenges, wheat crop disease is essential, potentially surpassing others in destructiveness. Timely diagnosis and recognition of wheat diseases are crucial for prevention and boosting national economies by increasing production yields. Conventional techniques like microscopic examination and manual visualization are time-consuming and prone to human errors. To overcome these limitations, researchers have introduced automatic techniques for identifying crop diseases [7]. These methods enable the agricultural industry to respond more effectively to crop diseases, ensuring sustained wheat production to meet the ever-growing global demand.

Several researchers attempted to develop machine learning (ML) and deep learning (DL) based networks to recognize diseases in wheat crops better. In the ML practices, the work [8] introduced a least squares regression model to identify early wheat disease severity with an overall accuracy of 82.35%. However, they just relied on conventional features resulting in poor performance. In the study [9], the authors developed an advanced ML system capable of recognizing major wheat diseases with good performance. However, they trained the mode on intra-class homogeneous data. Another research [10] focused on developing an image processing technique-based model for wheat disease recognition. In the study [11], au-

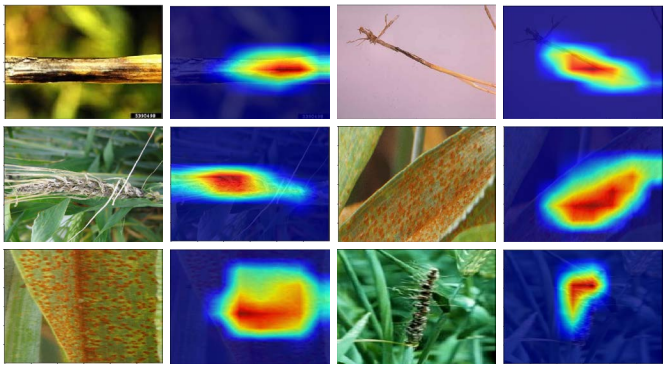


Fig. 1: Grad-CAM visualization for predicting a real case of wheat diseases

*Hyeonjoon Moon is a corresponding author.

thors proposed an approach for detecting buckwheat diseases, achieving an accuracy of 97.54% using a dual-layer inception structure and cosine similarity convolution. The authors in the study [12] applied the Segformer algorithm on stripe rust disease images, with improved performance through data augmentation. However, the limitation of this study regarding wheat disease lies in its specificity to fall wheat diseases in specific environments. The work [13] used ResNet-50 on edge devices for wheat yellow rust classification. However, the dataset utilized in this study lacks diverse images.

We take several key steps to bridge the research gaps to enhance the model's performance. Firstly, we enrich the dataset with more challenging samples to expose the model to complex data and improve its generalization. Next, we explore different CNN-based pre-trained models as backbones to identify optimal features for effective extraction. This step is crucial in improving the overall performance of the model. The emergence of attention mechanisms in deep learning plays a vital role in computer vision tasks [14]. This leads the model to notable improvements in tasks like object recognition. Thus, we introduced a modified soft attention mechanism to selectively boost the importance of relevant features before making the final prediction, refining the decision-making process. By combining these strategies, we aim to develop a robust and high-performing model capable of accurately addressing the complexities of the target task. The significant contributions of this work are elaborated as follows:

- Our study introduces new samples of challenging wheat field images by overcoming limitations in prior research using low-variance datasets. This practice enriches data diversity and complexity for the generalization of the model. A domain expert Integrated with benchmark-enhancing model training and evaluation labels the samples.
- The previous studies proposed pre-trained models without investigating the other CNN variants. We studied the impact of various models in our domain to find the optimal features. Resultantly, we find the InceptionResNetV2 as the best-performing model for feature extraction. Moreover, we investigated the model's different layers to take features from the higher representative layer.
- To enhance focus on intermediate features, we modified the soft attention module to focus on the challenging diseased areas. The progressive modification in the proposed model highlights the disease regions more preciously, ultimately leading the model to a comparatively better prediction of the baseline.
- We conducted extensive experiments on the proposed dataset and the training accuracy 97.02% and validation accuracy 95.66% shows the robustness of the model on challenging data.

The remaining part of the paper is structured as follows. Subsequently, Section 2 presents the suggested techniques, followed by findings and discourse in Section 3. Lastly, Section 4 covers the conclusion and potential areas for future research.

Fig. 2: Sample images of our dataset.

II. Proposed Methodolgy

This section represents the data collection details followed by the optimal feature extraction in our domain. In the end, the proposed model is comprehensively discussed.

A. Dataset

The dataset utilized in this paper comprises four distinct classes, encompassing three disease classes and one healthy class. Some of the data originates from the LWDCD2020 dataset. Our ultimate goal is to collect diverse data. The dataset employed in this study originates from wheat fields in the Khyber Pakhtunkhwa region, a notable wheat-producing area in Pakistan. The co-author, possessing considerable expertise in plant pathology, played a pivotal role in overseeing the data collection process. This expertise ensured the precise identification and selection of pertinent and characteristic samples, aligning with the local context. Our data gathering encompassed methodical visits to wheat fields across diverse locations within Khyber Pakhtunkhwa, thereby capturing varying environmental conditions and disease prevalence scenarios. The collected images encompass various viewpoints, intricate backgrounds, disease manifestations at different stages, and similar attributes shared among different wheat diseases. Figure 2 visualizes the distribution of each category. Among these, there are 1,436 images in the "healthy" class, 1,575 images in the "leaf rust" class, 910 images in the "crown and root rot" class, and 922 images in the "wheat loose smut" class. These images are allocated for training, and testing purposes, with the training, set constituting 80% of the entire dataset and the testing set making up 20% of the images. The overall statistics of the dataset are listed in Table I.

TABLE I: Statistics of our dataset

Classes	Training	Testing	Total
Healthy	1126	310	1436
Leaf rust	1260	315	1575
Crown and root rot	725	185	910
Wheat loose smut	763	159	922

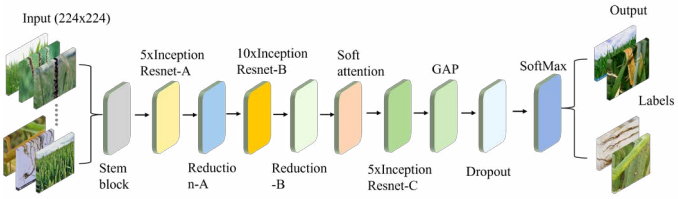


Fig. 3: The proposed framework for efficient wheat disease detection.

B. Optimal Features Extraction

Convolutional Neural Networks (CNNs) provide remarkable advantages for image analysis. These networks learn increasingly complex features, capturing fine details like edges and textures [15]. This hierarchical learning enables accurate image classification and diseased recognition. CNNs are also translation invariant, identifying patterns regardless of location and enhancing robustness. Their efficient techniques, like pooling, streamline processing without sacrificing essential information. CNNs-based networks are utilized for the better recognition of visual data in medicine [15], [16], disasters recognition [17] energy analytics [18], [19], anomaly detection [20], vehicle detection [21] and to improve the accuracy of models [**?**]. CNNs stand out as versatile and transformative networks for intricate [22] visual tasks. We conducted an extensive backbone models study to find the best-performing model. Resultantly, we find InceptionResNetV2 as the best-performing feature descriptor. Initially, the intermediate layer from each model was chosen for feature extraction. In this way, the final prediction layers were removed to evaluate the performance of the attention module. Moreover, we empirically identify the higher representative layer from the proposed backbone. This strategy enabled the utilization of well-established features while enhancing the overall performance of the models.

In the architecture of InceptionResNetV2 [23], the soft attention module is incorporated into the Inception Resnet C block of the model. This addition occurs in the part of the model where the image features are represented as an 8 x 8 feature size. The output of this max pool layer is then combined with the filter concatenate layer of the inception block. After the concatenate layer, a ReLU activation unit is applied. To regularize the output of the attention layer and prevent overfitting, a dropout layer with a rate of 0.5 is added following the activation unit. A batch normalization layer is incorporated after each layer in all the networks to introduce regularization. This helps stabilize the training process and improve the models' overall performance. In the case of the wheat dataset, which consists of 4 classes of wheat disease, the output layer is designed with four hidden units, followed by a softmax activation function. This configuration enables the model to perform multi-class classification. The network is trained for 10 epochs with a 0.01 and 32 batch size learning rate. The complete architecture of the network is depicted in Figure 3.

C. Attention Module

We address the challenge of focusing on wheat disease regions by integrating a soft attention mechanism into the established InceptionResNetV2 architecture. This strategy is rooted in the observation that intermediate features require heightened attention. To address this, we employ the soft attention mechanism, which selectively emphasizes relevant image regions while attenuating less crucial areas. This approach recognizes that certain image regions, such as the background, have minimal impact on classification accuracy. Conventional CNNs allocate uniform computational resources across the entire image, disregarding the varying significance of distinct regions. However, our adapted soft attention mechanism introduces an attention gate, strategically positioned 3D convolutional layers, and max pooling operations. This dynamic attention gate dynamically adjusts the importance of features, enabling the model to focus on critical patterns associated with different wheat classes, especially the nuanced distinctions between diseased and healthy crops. The attention mechanism allows us to allocate effectively on regions enhancing the model's ability to capture essential features. This streamlines processing by reducing computational redundancy and improves accuracy by concentrating on discriminative aspects of the input images. Intuitively, We extended the soft attention module by adding an extra 3D layer, enriching the model's capacity to grasp intricate relationships across multiple dimensions. This practice significantly enhances feature selection, culminating in more refined predictions. Furthermore, we integrated a 2x2 max pooling operation with 'same' padding into the concatenated features, enabling efficient downsampling while preserving spatial attributes. This selective process aids feature extraction and retains spatial structure. The use of 'same' padding ensures dimensional consistency, aligning with our strategic feature engineering approach.

$$f(s\beta) = \Upsilon t((\sum_{k=1}^{K} softmax(\mathbf{W}_k \times s))) \quad (1)$$

The feature tensor $t \in \Upsilon^{h \times w \times d}$ serves as the input to a 3D convolution layer with weights $\omega_\mu \in^{h \times w \times d \times \chi}$, where χ denotes the number of 3D weights. The outcome of this convolution is subjected to softmax, generating χ = 16 attention maps. These attention maps are then combined to form a unified attention map, acting as a weighting function denoted by β. The tensor t is then attentively scaled using α, further

adjusted by a learnable scalar Υ. The resulting attentively scaled features (fsβ) are subsequently concatenated with the original feature tensor t as a residual branch. During training, Υ is initialized from 0.01 to enable the network to gradually adapt and control the level of attention required for optimal performance.

III. Results

This study investigates the use of soft attention mechanisms in conjunction with seven deep CNN-based pre-trained models for wheat disease recognition. The implemented networks include MobileNetV2, VGG19, DenseNet121, ResNet101, EfficientNetB2, Xception, and InceptionResNetv2 [23], all of which are state-of-the-art feature extractors trained on the ImageNet dataset. Moreover, we empirically validated the impact of attention modules from different stages.

A. Implementation Details

The experiments were conducted on a computer system with an Intel(R) Core(TM) processor running at a clock speed of 3.70 GHz and 8.00 GB of RAM. The system also featured an RTX GFORCE 2070 GPU with 8 GB of onboard memory. The operating system used for the experiments was Microsoft Windows 10.

For the implementation of the project, various libraries were utilized. The Keras deep learning framework was employed with TensorFlow as the backend using Python version 3.7.16. Additionally, OpenCV version 4.8.0, a powerful computer vision library, is used for image preprocessing. The scikit-learn ML library version 1.0.2 was employed for training and testing various machine learning models. Furthermore, Matplotlib, a Python-based visualization library, generates visualizations for images, results, and graphs.

B. Feature Studies

The results of our study illuminate the comparative performance of various backbone architectures when incorporated into the MSA model. While all architectures demonstrate strong capabilities, InceptionResNetV2 stands out by achieving an outstanding accuracy of 95.66%. This remarkable performance positions InceptionResNetV2 as a compelling choice for tasks demanding high accuracy rates. In comparison, MobileNetV2 achieves an accuracy of 86%, VGG19 attains 88.4%, DenseNet121 reaches 89%, ResNet101 achieves 90%, EfficientNetB2 reaches 91.2%, and Xception attains 92% accuracy within the MSA framework. The dominance of InceptionResNetV2 over these renowned architectures underscores its ability to capture complex features and excel in challenging recognition tasks.

C. Impact of Attention

Modifications of networks in deep learning [24] are essential to enhance model performance and adapt to specific tasks [15], [25]. We evaluate the transformative impact of the modified soft attention mechanism when integrated with the InceptionResNetV2 backbone, and we contrast its effects with other backbone architectures, both with and without the attention mechanism. The practice of the attention mechanism yielded a discernible improvement in the model's performance, affirming its capacity to emphasize crucial features and suppress extraneous information selectively. Notably, the attention-enhanced InceptionResNetV2 achieved an accuracy of 95.66%, surpassing the baseline model's accuracy of 94% by 1.66% in accuracy. This substantiates the attention mechanism's potential to augment the feature extraction process, culminating in notable accuracy enhancements.

TABLE II: Features studies of our proposed method against another backbone method with SA and MSA implementation. The symbol ✓ represents that the attention module is applied and × shows that the attention is not applied

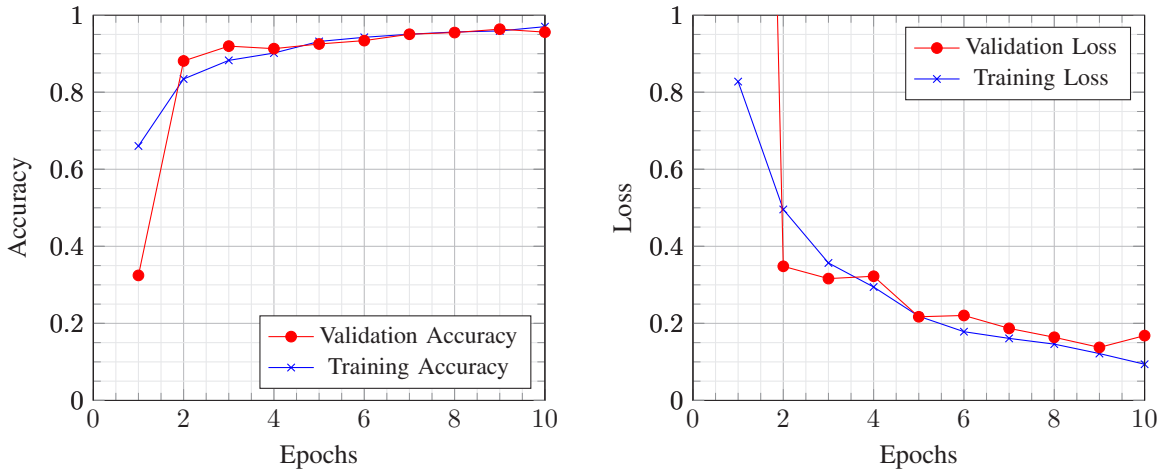
Backbone	SA	MSA	Accuracy (%)
MobileNetV2	×	✓	86
VGG19	×	✓	88.4
DenseNet121	✓	×	89
ResNet101	×	✓	90
EfficientNetB2	✓	×	91.2
Xception	✓	×	92
InceptionResNetV2	✓	×	94
InceptionResNetV2	×	✓	95.66

Furthermore, our analysis extended to the broader context of different backbone architectures. Across all architectures, incorporating the attention mechanism consistently translated to accuracy gains. This enhancement was particularly pronounced when coupled with the InceptionResNetV2 architecture. This observation underscores the complementary nature of the attention mechanism and the unique characteristics of InceptionResNetV2, collectively yielding a substantial accuracy boost. These findings illuminate the intricate interplay between architecture and attention mechanisms, highlighting the potential for specific backbones to synergize more effectively with attention-based enhancements. This subsection collectively underscores the attention mechanism's remarkable potential to elevate neural network performance, especially when harmonized with specific architectural traits, as demonstrated by the InceptionResNetV2 backbone. Figure 4 shows the accuracy and loss while Figure 5 represents the Confusion matrix of our proposed method.

D. Qualitative Results

To gain deeper insights into the contributions of the proposed modified soft attention mechanism, we employed the technique of explainable AI Gradient-weighted Class Activation Mapping (Grad-CAM) visualization. Grad-CAM offers a window into the decision-making process of neural networks by highlighting regions of input images that significantly influence the model's predictions. Our visualizations provide a compelling narrative of how the attention mechanism enhances the interpretability and effectiveness of the model.

The Grad-CAM visualizations unveiled that the attention mechanism consistently directs the network's focus to intricate and discriminative features. By emphasizing these



(a) Accuracy (b) Loss

Fig. 4: Validation, training accuracy and loss of our proposed method

(a) Before Modification (b) After Modificaiton

Fig. 5: Confusion Matrices of our proposed method before and after modification.

salient regions, the attention-enhanced model demonstrates a refined ability to capture intricate patterns, textures, and shapes, contributing to accurate predictions. This visualization-driven analysis not only corroborates the quantitative accuracy improvements but also offers a qualitative perspective on the attention mechanism's influence on the model's decision-making process. The Grad-CAM outcomes provide an intuitive and insightful means to comprehend the mechanism's efficacy and showcase its potential as shown in Figure 1

IV. Conclusion

In this paper, we introduced more challenging data related to wheat diseases to train a generalized model. We conducted extensive studies to find the optimal features specifically in our domain. To strengthen the intermediate features, we progressively modified the soft attention mechanism to focus deeply on diseased spots. We explore how the Soft Attention mechanism can be utilized to enhance the classification accuracy of wheat disease images with high-resolution content. Our model demonstrated superior performance compared to the current state-of-the-art methods on integrated datasets for wheat disease classification. Furthermore, we visualized the model attention on diseased spots using the technique of explainable AI called Grad-CAM. The results highlight the effectiveness of our approach in handling diverse wheat disease images, underscoring its potential for real-world applications in agricultural settings. In the future, this approach can be integrated into wheat disease IoT-supported systems. The proposed model will be deployed on resource-constraint devices to support agricultural experts and assist in crop disease diagnosis and management.

V. Acknowledgements

This work was supported by Seoul Campus Town Technology R&D Project (Development of a Genuine authentication solution to block counterfeit attempts using Artificial Intelligence (AI)-based deep learning technology) and by National Research Foundation of Korea (NRF) grant funded by the Korea government, Ministry of Science and ICT (MSIT) (2021R1F1A1046339) and by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries(IPET) through Digital Breeding Transformation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA)(322063-03-1-SB010).

References

[1] N. Poole, J. Donovan, and O. Erenstein, "Agri-nutrition research: Revisiting the contribution of maize and wheat to human nutrition and health," *Food Policy*, vol. 100, p. 101976, 2021.

[2] S. C. Bhardwaj, G. P. Singh, O. P. Gangwar, P. Prasad, and S. Kumar, "Status of wheat rust research and progress in rust management-indian context," *Agronomy*, vol. 9, no. 12, p. 892, 2019.

[3] L. M. Dang, M. Nadeem, T. N. Nguyen, H. Y. Park, O. N. Lee, H.-K. Song, and H. Moon, "Vpbr: An automatic and low-cost vision-based biophysical properties recognition pipeline for pumpkin," *Plants*, vol. 12, no. 14, p. 2647, 2023.

[4] L. M. Dang, K. Min, T. N. Nguyen, H. Y. Park, O. N. Lee, H.-K. Song, and H. Moon, "Vision-based white radish phenotypic trait measurement with smartphone imagery," *Agronomy*, vol. 13, no. 6, p. 1630, 2023.

[5] L. M. Dang, J. Shin, Y. Li, L. Tightiz, T. N. Nguyen, H.-K. Song, and H. Moon, "Toward explainable heat load patterns prediction for district heating," *Scientific Reports*, vol. 13, no. 1, p. 7434, 2023.

[6] D. Zhang, Q. Wang, F. Lin, X. Yin, C. Gu, and H. Qiao, "Development and evaluation of a new spectral disease index to detect wheat fusarium head blight using hyperspectral imaging," *Sensors*, vol. 20, no. 8, p. 2260, 2020.

[7] J. Boulent, S. Foucher, J. Théau, and P.-L. St-Charles, "Convolutional neural networks for the automatic identification of plant diseases," *Frontiers in plant science*, vol. 10, p. 941, 2019.

[8] I. H. Khan, H. Liu, W. Li, A. Cao, X. Wang, H. Liu, T. Cheng, Y. Tian, Y. Zhu, W. Cao *et al.*, "Early detection of powdery mildew disease and accurate quantification of its severity using hyperspectral images in wheat," *Remote Sensing*, vol. 13, no. 18, p. 3612, 2021.

[9] H. Khan, I. U. Haq, M. Munsif, Mustaqeem, S. U. Khan, and M. Y. Lee, "Automated wheat diseases classification framework using advanced machine learning technique," *Agriculture*, vol. 12, no. 8, p. 1226, 2022.

[10] P. Xu, G. Wu, Y. Guo, H. Yang, R. Zhang *et al.*, "Automatic wheat leaf rust detection and grading diagnosis via embedded image processing system," *Procedia Computer Science*, vol. 107, pp. 836–841, 2017.

[11] X. Liu, S. Zhou, S. Chen, Z. Yi, H. Pan, and R. Yao, "Buckwheat disease recognition based on convolution neural network," *Applied Sciences*, vol. 12, no. 9, p. 4795, 2022.

[12] J. Deng, X. Lv, L. Yang, B. Zhao, C. Zhou, Z. Yang, J. Jiang, N. Ning, J. Zhang, J. Shi *et al.*, "Assessing macro disease index of wheat stripe rust based on segformer with complex background in the field," *Sensors*, vol. 22, no. 15, p. 5676, 2022.

[13] U. Shafi, R. Mumtaz, M. D. M. Qureshi, Z. Mahmood, S. K. Tanveer, I. U. Haq, and S. M. H. Zaidi, "Embedded ai for wheat yellow rust infection type classification," *IEEE Access*, vol. 11, pp. 23 726–23 738, 2023.

[14] H. Khan, T. Hussain, S. Ullah Khan, Z. Ahmad Khan, and S. Wook Baik, "Deep multi-scale pyramidal features network for supervised video summarization," *Expert Systems with Applications*, p. 121288, 2023.

[15] H. Khan, M. Ullah, F. Al-Machot, F. A. Cheikh, and M. Sajjad, "Deep learning based speech emotion recognition for parkinson patient," *Image*, vol. 298, p. 2, 2023.

[16] M. Munsif, M. Ullah, B. Ahmad, M. Sajjad, and F. A. Cheikh, "Monitoring neurological disorder patients via deep learning based facial expressions analysis," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2022, pp. 412–423.

[17] M. Munsif, H. Afridi, M. Ullah, S. D. Khan, F. A. Cheikh, and M. Sajjad, "A lightweight convolution neural network for automatic disasters recognition," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2022, pp. 1–6.

[18] M. Munsif, F. U. M. Ullah, S. U. Khan, N. Khan, and S. W. Baik, "Ct-net: A novel convolutional transformer-based network for short-term solar energy forecasting using climatic information."

[19] M. Munsif, H. Khan, Z. A. Khan, A. Hussain, F. U. M. Ullah, M. Y. Lee, and S. W. Baik, "Pv-anet: Attention-based network for short-term photovoltaic power forecasting," pp. 133–135, 2022.

[20] S. Ul Amin, Y. Kim, I. Sami, S. Park, and S. Seo, "An efficient attention-based strategy for anomaly detection in surveillance video." *Computer Systems Science & Engineering*, vol. 46, no. 3, 2023.

[21] H. Khan, Z. A. Khan, W. Ullah, M. J. Kim, M. Y. Lee, and S. W. Baik, " ," , pp. 104–107, 2023.

[22] M. Ullah, S. U. Amin, M. Munsif, U. Safaev, H. Khan, S. Khan, and H. Ullah, "Serious games in science education. a systematic literature review," *Virtual Reality & Intelligent Hardware*, vol. 4, no. 3, pp. 189–209, 2022.

[23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[24] H. Khan, B. Q. Huy, Z. U. Abidin, J. Yoo, M. Lee, K. W. Seo, D. Y. Hwang, M. Y. Lee, and J. K. Suhr, "A modified yolov4 network with medium-scale challenging benchmark for efficient animal detection," , pp. 183–186, 2023.

[25] B. K. Yousafzai, S. A. Khan, T. Rahman, I. Khan, I. Ullah, A. Ur Rehman, M. Baz, H. Hamam, and O. Cheikhrouhou, "Student-performulator: Student academic performance using hybrid deep neural network," *Sustainability*, vol. 13, no. 17, p. 9775, 2021.