International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.4 (2023) pp. 325-330
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

A Novel Deep Learning Model for Pain Intensity Evaluation

Mahmut Emin ÇELİK [1]*

[1]Electrical Electronics Engineering Department, Faculty of Engineering, Gazi University, Ankara, Turkey
* **Corresponding Author** : **Email:** mahmutemincelik@gazi.edu.tr - **ORCID:** 0000-0002-1766-5514

Article Info:

DOI: 10.22399/ijcesen.1372628
Received : 07 October 2023
Accepted : 20 October 2023

Keywords

Pain intensity
Deep learning
Facial expression
Classification
Automated recognition

Abstract:

Pain assessment is a critical component of healthcare, influencing effective pain management, individualized care, identification of underlying issues, and patient satisfaction. However, the subjectivity and limitations of self-reported assessments have led to disparities in pain evaluation, particularly in vulnerable populations such as children, the elderly, individuals with cognitive impairments, and those with mental health conditions. Recent advances in technology and artificial intelligence (AI) have paved the way for innovative solutions in pain intensity evaluation.This paper presents a novel deep learning model to automatically classify pain intensity levels and compares them with six state-of-the-art deep learning classification models - ResNet-50, VGG-19, EfficientNet, DenseNets, Inception, and Xception- using the UNBC-McMaster Shoulder Pain Expression Archive Database for training. Transfer learning is employed to optimize model efficiency and minimize the need for extensive labeled data. Model evaluations are conducted based on accuracy, precision, recall, and F1 score. The proposed model, ZNet, showed superior performance of 95.4%, 64.4% and 63.4%, 63.7% for accuracy, precision, recall and F1-score respectively. Furthermore, this study addresses the challenge of accurately evaluating pain intensity in patients who cannot communicate verbally or face language barriers. By harnessing AI technology and facial expression analysis methods, we aim to provide an objective, reliable, and precise pain assessment methodology. Automated artificial based solutions enhance the reliability of pain evaluations, and holds promise for improving decision-making in pain management and treatment processes, ultimately enhancing patients' quality of life.

1. Introduction

Pain assessment is the process of evaluating and measuring an individual's pain experience to understand its nature, severity, and impact on their well-being. It is an essential component of healthcare, as it helps healthcare providers and clinicians make informed decisions regarding pain management and treatment. It is critically needed for the following reasons:

- Effective Pain Management: Accurate pain assessment is crucial for providing appropriate pain relief. It helps healthcare professionals determine the most suitable interventions, medications, or therapies to alleviate pain and improve the patient's comfort and quality of life.
- Individualized Care: Pain experiences can vary greatly from person to person. Pain assessment allows healthcare providers to tailor their approach to each patient's unique needs, ensuring that treatment plans are individualized and effective.
- Identification of Underlying Issues: Pain can be a symptom of an underlying medical condition. By assessing pain comprehensively, healthcare providers can identify potential causes that may require further investigation and treatment.
- Patient Satisfaction: Adequate pain management and communication about pain contribute to higher levels of patient satisfaction with their healthcare experience.

There is a notable absence of an effective and dependable method for the objective quantification of an individual's pain experience. Healthcare professionals and organizations predominantly depend on a patient's self-reported assessment to ascertain the severity of pain; however, these approaches may exhibit limitations and potential inaccuracies. On the other hand, there are several

disadvantaged groups of patients that pain intensity evaluation is of great importance due to the impossibility of communication [1,2]. The first group is children and the elderly. Vulnerable populations like children and the elderly may experience disparities in pain assessment. Young children may have difficulty communicating their pain, while older adults may have their pain dismissed as a natural part of aging. The other is individuals with cognitive impairments. People with cognitive impairments, such as dementia or intellectual disabilities, may have difficulty expressing their pain or may not be taken seriously when they do communicate their pain. This can lead to underassessment and undertreatment. Lastly, individuals with mental health conditions may experience disparities in pain assessment. Pain complaints in these individuals can sometimes be attributed to their psychiatric condition, and their physical pain may be overlooked or undertreated. In recent years, advances in technology, along with multidisciplinary studies in medicine and engineering, have enabled the development of various approaches to understand and manage pain. The use of AI techniques for pain intensity evaluation contributes to the more effective assessment of patients and the creation of appropriate treatment plans [1-4]. Artificial intelligence (AI) plays a pioneering role in various fields of research including medicine. It has a wide range of applications providing a crucial support in medical and dental diagnosis and treatment processes [5-7]. Table 1 presents previous works using deep learning. AI-based automated solutions are now needed to reduce the subjectivity of existing methods that often rely on subjective and inconsistent reports from patients or clinicians due to lack of measurements and different conceptualizations. The present study aims:

- to develop a novel deep learning model for pain intensity evaluation that can automatically
- detect the pain level using facial expressions with 4 different levels.
- to compare the proposed model with the state of art six different deep learning classification models.

The publicly available dataset, UNBC-McMaster Shoulder Pain Expression Archive Database, is used for training the models. The transfer learning concept has been applied to leverage pre-existing knowledge, reduce the need for massive amounts of labeled data, and improve the efficiency of model training. Deep learning models are evaluated by accuracy, precision, recall and F1 score.

Table 1. *Comparison of previous studies with details*

Reference	Purpose	Dataset	Method	Results
Weitz et al [1]	Distinguishing expressions of pain from emotions such as happiness and disgust using facial expressions	Pain Intensity 3: 12,006 Pain Intensity 4: 12,006 Disgust: 24,075 Happiness: 24,075	VGG-Face LRP (Layer-wise Relevance Propagation)	Pain: Precision: 0.62 Recall: 0.69 F1-score: 0.66 Disgust: Precision: 0.70 Recall: 0.73 F1-score: 0.71 Happiness: Precision: 0.67 Recall: 0.57 F1-score: 0.62
Prabal Datta et al [18]	Assisting physicians in non-verbal identification of pain using facial images for uncommunicative patients or during surgery	UNBC: 10852 frames DISFA: 39182 frames	DarkNet19	UNBC-McMaster Acc: 95.57% UAR: 95.59% UAP: 95.79% Average F1: 95.67% MCC: 94.14% CK: 93.93% GM: 95.58% DISFA Acc: 96.06% UAR: 96.04% UAP: 96.16% Average F1: 96.08% MCC: 94.78% CK: 94.74% GM: 96.03%
Zakia Hammal and Jeffrey F. Cohn [19]	To reliably measure pain intensity Pain assessment and management	16,657	AAM (Active Appearance Model): are used to extrac the canonical appearence of the face (CAPP) Log Normal Filters SVM (for classification)	Mean CR: 97% Mean PR: 96.75% Mean F1-score: 93.5%
Fontaine et al	Pain intensity evaluation for especially non- communicating people in clinical condition	2810	ResNet- 18	Accuracy: 32% - 53%

This work contributed to the existing technology as follows:

- it proposes a new deep learning model, namely ZNet
- it compares proposed model with existing 6 classification models
- it provides superior performance with transfer learning
- it promises for improved decision-making
- it comprehensively discusses current limitations towards clinical integration.

2. Material and Methods

The facial expressions are based on the Prkachin and Solomon Pain Intensity (PSPI) metric that relies on the Facial Action Coding System (FACS) and is widely used for manual pain assessment [8]. The UNBC-McMaster Shoulder Pain Expression Archive Database is publicly available, and it is used for training the models [8]. The UNBC-McMaster database has been classified according to the PSPI metric, resulting in 16 different levels. To facilitate the analysis and balance the data, we performed class merging inspired by previous works, resulting in 4 classes [4,9,10]. Each class includes PSPI0, PSPI1-2, PSPI3-4, and PSPI4+ respectively. Each class, namely, class0, class1-2, class3-4 and class4+ has 40,029, 5,260, 2,214 and 895 frames respectively. Figure 1 demonstrates example images from the dataset for each class. Data is divided into training, validation and test set with the ratio of 70%, 20% and 10%. Additionally, a preprocessing step normalized and resized images to 128x128 pixels to get more stable training process. The state of art 6 different deep learning classification models is used to classify pain intensity using facial expressions. Namely, ResNet-50, VGG-19, EfficientNet, DenseNets, Inception, and Xception were implemented [11-16]. Additionally, a novel custom-designed deep learning model is proposed for the same problem. Model performances are evaluated by widely used evaluation metrics that might be considered as a standard for the classification tasks. Accuracy, Precision, Recall, and F1-score are employed to determine classification ability of the models. Briefly, accuracy metric measures the model's correct classification rate, while precision and recall metrics assess how the model classifies positive and negative classes. The F1-score combines precision and recall metrics to measure the overall performance of the model. These metrics serve as essential tools to understand how effective the models are in real-world applications and to enhance their performance. Figure 2 shows the flowchart of the deep learning application.

Figure 1. *Example images for each class form the dataset*

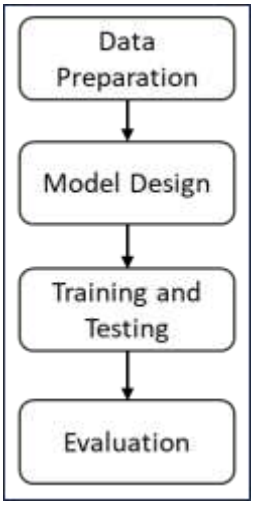

Figure 2. *Flowchart of the application*

2.1 Proposed Model: ZNet

ZNet model is inspired by the UNet architecture. The model architecture can be explained as follows. Figure 3 presents model architecture.

The ZNet model consists of double convolutional blocks, max-pooling layers, and Z-point merging layers for feature extraction and learning.

The configuration of the model is as follows:

•Input Layer: The model takes color images as input (n_channels=3).

•Double Convolutional Blocks (DoubleConv): Each DoubleConv block comprises two consecutive convolutional layers, batch normalization, and ReLU activation functions. These blocks serve as the basic building blocks for learning and feature extraction.

•Max Pooling Layers (MaxPoolLayer): Max pooling layers are used in the Znet model to reduce the size of feature maps. These layers significantly decrease the computation cost and the number of learnable parameters.
•Z-point Merging Layers (Zpoint): Zpoint layers are employed to merge feature maps with different dimensions. This allows the model to combine both low-level and high-level features for better feature extraction.
•Batch Normalization Layers: Batch normalization is a technique that accelerates the training of the network and reduces the risk of overfitting. In the ZNet model, batch normalization layers are used to stabilize the model's performance.
•Classification Layer (classifierPart): The classification layer at the output of the ZNet model is used for pain level classification. This layer flattens the feature maps and uses fully connected layers to predict the pain level.

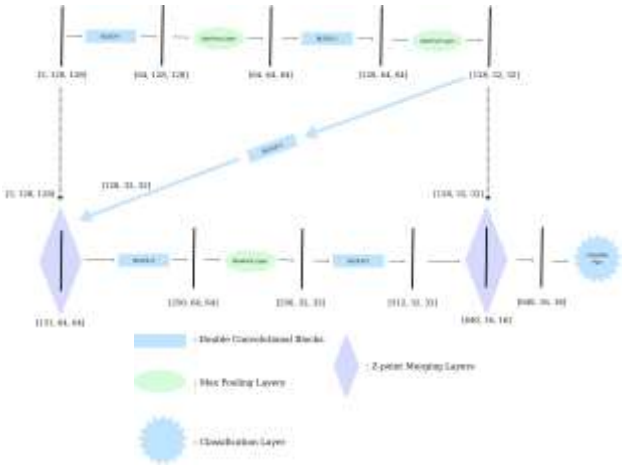

Figure 3. *ZNet architecture representation*

It is aimed to enhance model's performance by using skip connections to directly transmit low-level features to high-level features. These connections allow our model to transfer the knowledge of low-level features to more complex and meaningful features in the high-level layers. Skipped connections are a widely used technique, especially in models like UNet, and they often yield good results in segmentation and image processing tasks. Low-level features usually contain more basic and local information, while high-level features represent more abstract and comprehensive information. Therefore, skipped connections contribute to the model's ability to generalize better. However, each connection passing through each layer adds to the increase in the number of parameters. This is an important consideration in model training and performance evaluation [17]. Using skipped connections too frequently can lead to overfitting and unnecessary growth of weight matrices. Hence, it is needed to carefully determine the skip connections and reduce them when necessary. This way, we can achieve good performance while keeping the number of parameters under control, making the model training more effective.

3. Results and Discussions

Model performances are obtained and presented as follows. The confusion matrix for each model, Inception, Xception, Densenet, EfficientNet, Znet, ResNet-50, and VGG-19, are used to get evaluation metrics. Inception achieved a validation accuracy of 88.4% but struggled with precision and recall, scoring 48.9% and 45.5% respectively. The F1-score of 46.4% indicates a moderate balance between precision and recall. On the test set, Inception achieved a decent general accuracy of 87.7%, but the accuracy for certain classes (class1-2, class3-4, and class4+) was significantly lower. The Xception demonstrated better results than Inception, with a validation accuracy of 95.3% and higher precision and recall scores of 49.7% and 48.4% respectively. However, the F1-score remained relatively low at 48.9%. On the test set, Xception performed well overall with a general accuracy of 95.4%. However, particularly class3-4 and class4+ provided lower rates. The Densenet achieved a validation accuracy of 92.2% with balanced precision and recall scores of 59.3% and 57.3% respectively. The F1-score of 58.0% indicates a satisfactory balance. On the test set, Densenet demonstrated a good general accuracy of 91.9%, but similar to other models, it faced challenges with several other class accuracies. The EfficientNet achieved a validation accuracy of 89.4% with relatively low precision and recall scores of 48.8% and 45.9% respectively. The F1-score was 46.7%, indicating a need for better improvement in handling class imbalances. On the test set, EfficientNet showed a general accuracy of 89.5%. ResNet-50 achieved a validation accuracy of 92.0% with precision and recall scores of 42.8% and 40.8% respectively, resulting in an F1-score of 41.6%. On the test set, ResNet-50 demonstrated an overall accuracy of 91.6%, with better performance for class0 but lower accuracy for other classes. VGG-19 achieved a validation accuracy of 91.9% with balanced precision and recall scores of 66.4% and 64.1% respectively, leading to an F1-score of 64.8%. On the test set, VGG-19 showed high accuracy for class0 but decreased with several class accuracies, especially for class1-2 and class3-4. ZNet, a novel model inspired by skip connections and Unet architecture, demonstrated remarkable performance. It achieved a validation accuracy of 95.4%, with precision and recall scores of 64.4% and 63.4%,

respectively, resulting in an F1-score of 63.7%. The model exhibited a strong balance between precision and recall. On the test set, ZNet excelled in classifying minority classes (class1-2, class3-4, and class4+), achieving an accuracy of 85.2%, 85.5%, and 69.7% respectively. The overall test accuracy of 96.5% showcases ZNet's capability to achieve superior performance in real-world applications.
Overall, the proposed model achieves superior and successful results. In real-life tests conducted on the test dataset, we can observe that the ZNet model achieved better results compared to other models in terms of both class-specific and overall accuracy values, as shown in Table 2.

Table 2. *Class-based performance of the models for testing*

Model	**Accuracy (%)**				
	Class 0 ($PSPI_0$)	Class 1-2 ($PSPI_{1-2}$)	Class 3-4 ($PSPI_{3-4}$)	Class4 + ($PSPI_{4+}$)	Average
Densenet	98.13	67.30	51.13	59.55	91.92
Efficient Net	97.93	55.70	37.10	40.45	89.50
Inception	97.95	40.87	35.29	33.71	87.70
ResNet-50	96.75	65.97	70.14	61.80	91.55
VGG-19	**99.35**	55.89	57.01	53.93	91.86
Xception	97.48	**90.87**	*79.74*	*65.17*	*95.35*
ZNet	*99.20*	*85.17*	**85.52**	**69.66**	**96.51**

Automated pain level evaluation research using modern artificial intelligence technologies has an increasing trend in medical informatics. This work explores the applicability and performance of various deep learning models for pain level classification using facial expressions. Results indicate that the proposed model in addition to other deep learning models can successfully predict pain level classes with high performance. The proposed model yielded superior performance for three out of four classes in the dataset. Deep learning models have shown promising results in accurately categorizing pain intensity levels. However, it is important to acknowledge that the accuracy of these models may vary depending on the dataset, the choice of architecture, and the quality of the input data. Transitioning from research to practical clinical applications requires careful consideration. The integration of deep learning models into healthcare settings demands robust validation, collaboration with medical professionals, and adherence to regulatory guidelines. Additionally, the development of user-friendly interfaces for healthcare providers is essential to facilitate adoption. Previous studies including intensity level classification using deep learning are investigated for comparison purposes. Table 2 lists previous studies based on their purpose, dataset size, method and results reported. Weitz et al applied a deep learning model to distinguish pain expressions from emotions such as happiness and disgust using facial expressions [1]. Prabal Datta et al used DarkNet19 to assist physicians in non-verbal identification of pain using facial images for uncommunicative patients or during surgery [18]. Zakia Hammal et al used SVM for classification to reliably measure pain intensity pain assessment and management [19]. Fontaine et al applied ResNet-18 classification model to facial expressions of patients taken in clinical conditions, then compared them with the findings of nurses related to pain level of the patients [2]. There are several limitations, firstly the lack of a wide balanced public dataset of facial expressions. Previous studies have not used significant open datasets, making it difficult to compare the performance of deep learning models. In order to apply the developed models, it is necessary to use large, balanced datasets that have been identified by doctors. In the open-source dataset used in this study, the number of categories was reduced from 16 to 4 to ensure balance between classes as applied in previous studies, but a more balanced dataset is ideally an important requirement for clinical integration. Data privacy and ethical concerns should be considered, which may limit the ability to prepare and share a large multi-centered data set.

4. Conclusions

The application of deep learning techniques for the evaluation of pain intensity levels represents a significant advancement in the field of healthcare and pain management. Deep learning models have demonstrated their effectiveness in categorizing pain intensity levels. In conclusion, the use of deep learning for pain intensity level evaluation is a promising development with the potential to revolutionize pain management and improve the lives of countless individuals. As researchers and healthcare providers continue to work on refining and implementing these technologies.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have

appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]. Weitz, K., Hassan, T., Schmid, U., & Garbas, J. U. (2019). Deep-learned faces of pain and emotions: Elucidating the differences of facial expressions with the help of explainable AI methods. *tm-Technisches Messen*, 86(7-8); 404-412. DOI: 10.1515/teme-2019-0024.

[2]. Fontaine, D., Vielzeuf, V., Genestier, P., Limeux, P., Santucci-Sivilotto, S., Mory, E., ... & DEFI study group. (2022). Artificial intelligence to evaluate postoperative pain based on facial expression recognition. *European Journal of Pain*, 26(6); 1282-1291. DOI: 10.1002/ejp.1948.

[3]. Hasan, M. K., Ahsan, G. M. T., Ahamed, S. I., Love, R., & Salim, R. (2016). Pain level detection from facial image captured by smartphone. *Journal of Information Processing*, 24(4); 598-608. DOI: 10.2197/ipsjjip.24.598.

[4]. Barua, P. D., Baygin, N., Dogan, S., Baygin, M., Arunkumar, N., Fujita, H., ... & Acharya, U. R. (2022). Automated detection of pain levels using deep feature extraction from shutter blinds-based dynamic-sized horizontal patches with facial images. *Scientific reports*, 12(1); 17297. DOI: 10.1038/s41598-022-21380-4.

[5]. Çelik, B., & Çelik, M. E. (2023). Root Dilaceration Using Deep Learning: A Diagnostic Approach. *Applied Sciences*, 13(14); 8260. DOI: 10.3390/app13148260.

[6]. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., & Nandi, A. K. (2022). Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5); 1243-1267. DOI: 10.1049/ipr2.12419.

[7]. Çelik, B., & Çelik, M. E. (2022). Automated detection of dental restorations using deep learning on panoramic radiographs. *Dentomaxillofacial Radiology*, 51(8); 20220244. DOI: 10.1259/dmfr.20220244.

[8]. Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., & Matthews, I. (2011, March). Painful data: The UNBC-McMaster shoulder pain expression archive database. In 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG) (pp. 57-64). IEEE. DOI: 10.1109/FG.2011.5771462.

[9]. Bargshady, G., Zhou, X., Deo, R. C., Soar, J., Whittaker, F., & Wang, H. (2020). Enhanced deep learning algorithm development to detect pain intensity from facial expression images. *Expert Systems with Applications*, 149; 113305. DOI: 10.1016/j.eswa.2020.113305.

[10]. Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., ... & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. *ACM Computing Surveys (CSUR*), 55(3); 1-37. DOI: 10.48550/arXiv.2111.08834.

[11]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). DOI: 10.48550/arXiv.1512.03385.

[12]. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. DOI: 10.48550/arXiv.1409.1556.

[13]. Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR. DOI: 10.48550/arXiv.1905.11946.

[14]. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). DOI: 10.48550/arXiv.1608.06993.

[15]. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9). DOI: 10.48550/arXiv.1409.4842.

[16]. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258). DOI: 10.1109/CVPR.2017.195.

[17]. Erol, T., & Sarikaya, D. (2022). PlutoNet: An Efficient Polyp Segmentation Network with Modified Partial Decoder and Decoder Consistency Training. *arXiv preprint arXiv*:2204.03652. DOI: 10.48550/arXiv.2204.03652.

[18]. Barua, P. D., Baygin, N., Dogan, S., Baygin, M., Arunkumar, N., Fujita, H., ... & Acharya, U. R. (2022). Automated detection of pain levels using deep feature extraction from shutter blinds-based dynamic-sized horizontal patches with facial images. Scientific reports, 12(1), 17297. DOI: 10.1038/s41598-022-21380-4.

[19]. Hammal, Z., & Cohn, J. F. (2012, October). Automatic detection of pain intensity. In Proceedings of the 14th ACM international conference on Multimodal interaction (pp. 47-52). DOI: 10.1145/2388676.2388688.