DOI: 10.22146/jnteti.v13i2.9466

Learning Rate Analysis for Pain Recognition Through Viola-Jones and Deep Learning Methods

Raihan Islamadina[1,2], Khairun Saddami[3], Fitri Arnia[3], Taufik Fuadi Abidin[3], Rusdha Muharar[3], Muhammad Irwandi[2], Aulia Syarif Aziz[2]

[1] Doctoral Program, School of Engineering, Faculty of Engineering, Syiah Kuala University, Banda Aceh, Aceh 23111, Indonesia
[2] Prodi Pendidikan Teknologi Informasi, Fakultas Tarbiyah dan Keguruan, Universitas Islam Negeri Ar-Raniry, Banda Aceh, Aceh 23111, Indonesia
[3] Department of Electrical and Computer Engineering, Faculty of Engineering, Syiah Kuala University, Banda Aceh, Aceh 23111, Indonesia

[Accepted: 30 August 2023, Revised: 1 November 2023, Accepted: 14 March 2024]
Corresponding Author: Fitri Arnia (email: f.arnia@unsyiah.ac.id)

ABSTRACT — Deep learning is growing and widely used in various fields of life. One of which is the recognition of pain through facial expressions for patients with communication difficulties. Viola-Jones is a simple algorithm that has real-time detection capabilities with relatively high accuracy and low computational power requirements. The learning rate is a significant number that has an impact on the deep learning result. This study recognized pain using the Viola-Jones and deep learning methods. The dataset used was a thermal image from the Multimodal Intensity Pain (MIntPAIN) database. The steps taken consisted of segmentation, training, and testing. Segmentation was conducted using the Viola-Jones method to get the significant area of the face image. The training process was carried out using four deep learning benchmarks model, which were DenseNet201, MobileNetV2, ResNet101, and EfficientNetb0. Besides that, deep learning has a very important number to determine that is learning rate, which impact the deep learning results. There were five learning rates, which were 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. Learning rate values were then compared with four deep models learning to obtain high accuracy results in a short time and simple algorithm. Finally, the testing process was carried out on test data using a deep learning benchmark model in accordance with the training process. The research results showed that a learning rate of 10^{-2} from the MobileNetV2 method produced an optimal performance with a training validation accuracy of 99.60% within a time of 312 min and 28 s.

KEYWORDS — Pain Recognition, Viola-Jones Method, Learning Rate, Deep Learning, Accuracy.

I. INTRODUCTION

The innate human tendency to convey inner emotions finds its manifestation in the form of facial expressions, which serve as a powerful medium of nonverbal communication. This phenomenon of utilizing facial cues for conveying emotions holds significant relevance in the context of evaluating pain [1]. Traditionally, the assessment of pain has heavily relied upon an individual's own verbal account. Nevertheless, situations arise, as in the case of individuals grappling with terminal illnesses and rendered incapable of conventional communication, where self-reported pain becomes a complex puzzle to decipher, potentially leading to misinterpretation and erroneous conclusions.

The intricate task of identifying pain through analyzing facial expressions stands as a multifaceted challenge necessitating the development and implementation of resilient methodologies. The process of facial recognition encompasses three pivotal phases, namely the initial detection of faces, subsequent extraction of distinctive features, and ultimately, the recognition of the facial attributes. In the realm of face detection, the algorithm proposed by Viola-Jones emerges as the preeminent and extensively adopted solution [2], tracing its origins back to its introduction in the year 2001 [3].

Renowned for its efficacy, the Viola-Jones algorithm has garnered widespread acclaim primarily due to its straightforwardness and its remarkable proficiency in real-time face detection. This approach boasts a commendable balance between accuracy and the demand for computational resources, thereby making it a preferred choice across various applications [4]–[6]. The algorithm's inherent capacity for swift and accurate detection has positioned it as an indispensable tool, particularly well-suited for scenarios where timely responsiveness and efficiency are imperative.

The exploration of pain recognition has witnessed a triumphant convergence with the realm of deep learning methodologies, marking a significant advancement in this domain [7]. The deep learning was utilized in the extraction of intricate features, thereby facilitating the detection of pain through the analysis of facial expressions [8], [9]. Furthermore, another notable study has harnessed the power of deep learning to navigate the landscape of pain recognition, specifically focusing on self-reported pain levels using the visual analogue scale (VAS) [10].

The evolution of deep learning techniques in the realm of pain recognition continues its stride, leveraging an expansive RGB image dataset sourced from the Multimodal Intensity Pain (MIntPAIN) database. A noteworthy investigation yielded an impressive accuracy rate of 92.26% [11], This accomplishment is paralleled by the findings of another study, which secured an accuracy level of 92.44%, reaffirming the prowess of deep learning methodologies in tackling the intricate task of pain recognition [12]. Moreover, the horizon of deep learning's applications was broadened to encompass the utilization of thermal image datasets from the MIntPAIN database, as demonstrated by the endeavors of previous researchers [13]. Their efforts yielded a commendable accuracy rate of 83.5%, thus underscoring the versatility of deep learning approaches in the multifaceted domain of pain recognition.

The present study represents a significant evolution built upon the groundwork established by a preceding investigation [13], which focused on the identification of pain through the

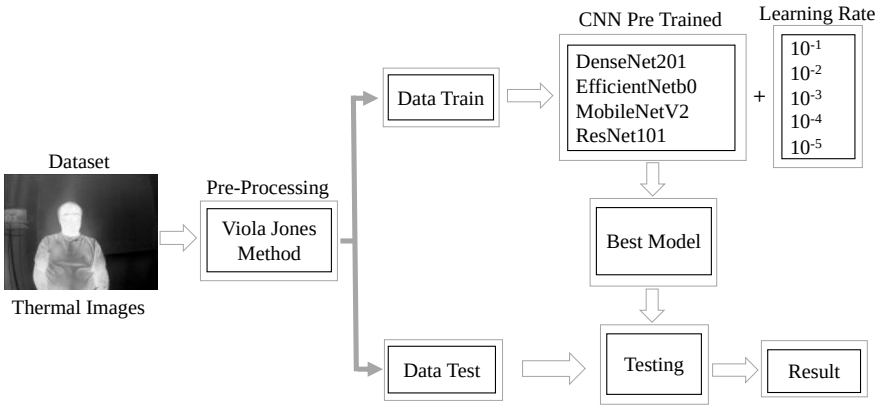

Figure 1. Research stages.

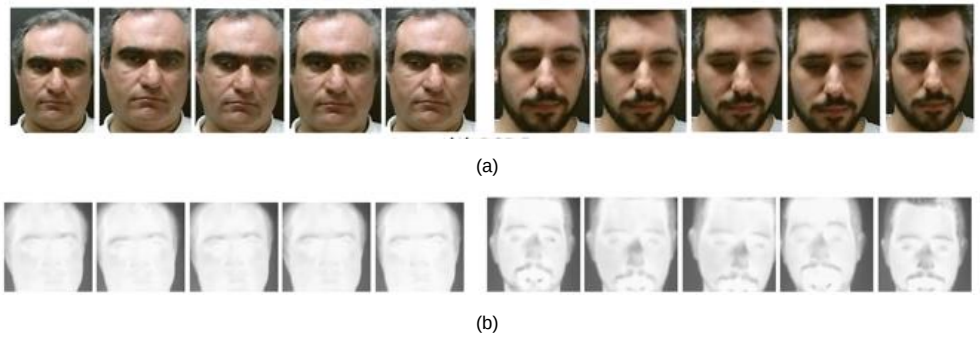

Figure 2. Pain level dataset, (a) RGB face images and (b) thermal face images.

combined utilization of the Viola-Jones and deep learning methodologies. In this progressive endeavor, the core dataset under scrutiny consists of thermal images extracted from the MIntPain database. The methodological trajectory unfolds through distinct stages, encompassing segmentation, training, and subsequent testing.

The initial stage, namely segmentation, drew upon the Viola-Jones method, effectively pinpointing crucial regions within facial images that bear relevance to the pain recognition task. This strategic step serves as the preliminary filtering mechanism, sifting through the visual data to identify salient features. Subsequently, the training phase was conducted, leveraging the prowess of four deep learning benchmarks as foundational pillars. These benchmarks encompass a diverse array of models, including the dense convolutional network model (DenseNet201) [14], MobileNetV2 [15], residual network (ResNet101) [16], and EfficientNetb0 [17].

Integral to the research's success is the determination of the learning rate, a hyperparameter pivotal in governing the extent of model adjustments in response to computed errors during weight updates [18]. In this endeavor, a range of five learning rates namely, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}, is employed, each playing a crucial role in shaping the outcomes of the ensuing deep learning processes [18].

In effect, the study endeavors to conduct a comprehensive comparison amongst the four distinct deep learning models, juxtaposing their performance with varying learning rate values. The primary objective here is twofold: to achieve heightened accuracy levels within a compressed time frame and to retain the inherent simplicity of the algorithmic methodologies employed.

The final phase of this study culminates in the rigorous testing of the evolved model. The testing procedure is meticulously executed on data specifically earmarked for validation, utilizing the very same deep learning benchmark models that were cultivated during the training phase. This congruence between training and testing models serves to authenticate the translational applicability of the developed framework, ensuring that the insights gleaned from the training process seamlessly manifest in real-world scenarios.

II. METHODOLOGY

The stages carried out in this study are described in Figure 1.

A. DATA COLLECTION

For the purpose of this study, the dataset selected for analysis originated from the MIntPAIN database, specifically in the format of thermal images. This database contains an extensive array of 9,366 variables derived from 20 subjects [19]. To facilitate a comprehensive evaluation, the dataset was systematically partitioned into two distinct subsets: the training and the test datasets.

The training dataset contained a total of 5,000 thermal images, while the test dataset was comprised of 1,600 images. The dataset collectively covered a spectrum of pain intensity, encompassing a total of five distinct pain levels, each ranging from level 0 indicating "no pain" to level 4 signifying a "high pain level", as shown in Figure 2. This categorization of pain levels offers a nuanced perspective that captures the diversity of pain experiences and aids in training and validating the models for effective pain recognition.

B. DATA PREPROCESSING

The data preprocessing phase in this study was conducted through the adept application of the Viola-Jones method, a technique employed to extract the facial regions within the images. This initial step involved the manipulation of input images, effectively isolating and extracting critical facial components that had been identified as pivotal for subsequent

analysis. The inception of this process was marked by the utilization of the Haar-like feature technique, a method that systematically breaks down the image into discrete regions, progressing from the top-left corner to the bottom-right corner. A fundamental augmentation to this technique is the incorporation of the integral image procedure, a computational strategy designed to expedite object detection [3].

An additional layer of sophistication was introduced through the amalgamation of multiple weak classifiers into a more potent classifier, achieved through the Adaptive Boosting technique. Adaptive Boosting serves as a mechanism to synergistically harness the strengths of these weak classifiers, resulting in the formulation of a robust and adept classifier. This process began with the calculation of weight, which was then followed by the evaluation of feature values for each weak classifier [3]. Subsequent decisions hinge upon the assessment of these feature values; if the value fell below 0, the image was categorized as devoid of an object. Conversely, if the value exceeded 1, the image was deemed to host an object. The cascade classifier approach increased the complexity by integrating intricate classifiers within a hierarchical framework, resulting in faster object detection.

The culmination of this preprocessing stage was marked by the implementation of bounding box techniques, instrumental in demarcating and delineating the detected facial entities [3]–[5]. This process was conducted to isolate the critical facial regions. This strategy was subsequently applied to both the training and testing image datasets. The delineated facial regions were then subjected to cropping or truncation, resulting in the extraction of specific facial contours. This cropping strategy was essential for facilitating the subsequent identification of facial contours, effectively simplifying the overall process. This cropping procedure also triggered a transformation of the image dimensions, transitioning from the original 640 × 480 pixels to a more compact 104 × 104 pixels post-cropping. Visual representation of the outcomes of the bounding box and facial area cropping procedures can be observed in Figure 3, offering a tangible glimpse into the results attained at this juncture of the preprocessing pipeline.

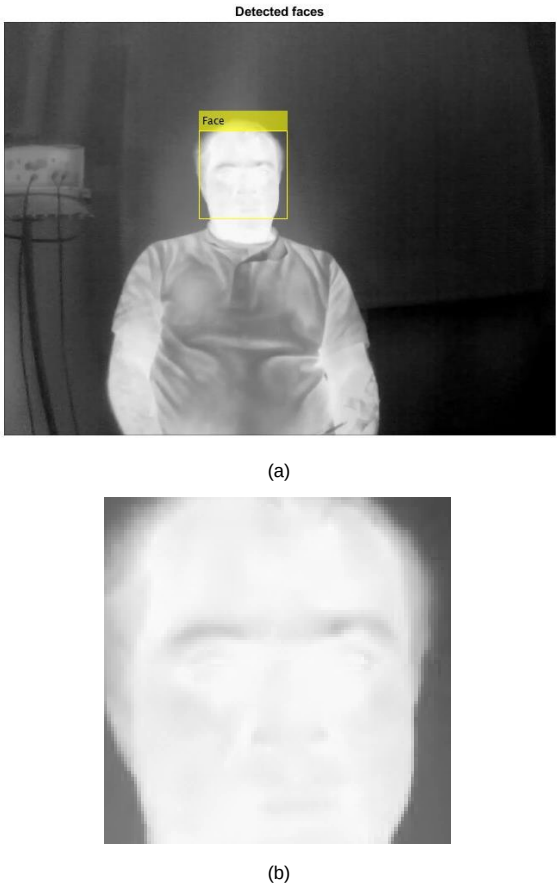

Figure 3. Visual representation of (a) the bounding box image of the face area and (b) the cropped area of the face.

C. DATA TRAINING

In this study, the training process was performed through the integration of pretrained convolutional neural network (CNN) models, which are leveraged in conjunction with transfer learning techniques. This approach was guided by the imperative of efficiency and expediency, utilizing well-established models such as DenseNet201, EfficientNetb0, MobileNetV2, and ResNet101.

DenseNet201 establishes connections between every layer in a feed-forward fashion. The working principle of DenseNet201 involves concatenating the output from the previous layer. Initially, a 28 × 28 × 3 image size was spread over 24 channels which resulted in an image size of 28 × 28 × 24. Furthermore, 12 features with the same width and height were used in each subsequent convolution layer, producing an output layer of 28 × 28 ×12. In the next layer, the input was 28 × 28 × 24 + 12, then, in the next layer, it was 28 × 28 × 24 + 12 + 12, and so on. In this way, DenseNet201 can reduce overfitting and use fewer parameters [14].

MobileNetV2 can improve the performance of mobile models on many tasks and benchmarks across a spectrum of different model sizes. MobileNetV2 is based on an inverted residual structure in which the bypass connections are between thin layers of the bottleneck. The MobileNet architecture is changed by replacing the fully convolutional operator with a factorization that divides the convolution into two layers: a depthwise convolution layer and a pointwise convolution layer [15].

ResNet101 is a pretrained model that can manage its layers without needing specific configurations. ResNet101 is used to overcome performance degradation on deep networks. This model can build deeper networks and can find the number of layers optimized for missing gradients. ResNet101 has 101 layers and is able to classify 1,000 image categories [16].

EfficientNetb0 scales all dimensions from depth, width, and resolution uniformly using compound coefficients to deliver better performance. EfficientNet is also able to transfer well and achieve the best accuracy on CIFAR-100 (91.7%), flower (98.8%), and three other transfer learning datasets with fewer parameters [17]. These models were equipped with preexisting knowledge and patterns extracted from the training on extensive datasets, rendering them as optimal starting points for further training.

Transfer learning, a central facet of this approach, operated on the premise of utilizing a pretrained model as a foundational scaffold and then refining its parameters through fine-tuning to align with the new dataset under consideration. This technique capitalizes on the wealth of knowledge encapsulated within the pretrained model. In addition, through iterative adjustments, it adapts to the specific features and intricacies of the target dataset. The primary goal of transfer learning is to capitalize on the preexisting capabilities of the model in its original domain and extend its utility to solve related problems in a different domain.

TABLE I
PRETRAINED NETWORK

Network	Depth	Size (MB)	Parameter (millions)	Image Input Size
DenseNet201	201	77	20.0	224-by-224
EfficientNetb0	82	20	5.3	224-by-224
MobileNetV2	53	13	3.5	224-by-224
ResNet101	101	167	44.6	224-by-224

The pretrained model serves as a feature extractor, distilling high-level image representations that hold significance across various domains. By adjusting the parameters of the pretrained model in harmony with the new dataset, the model's capabilities were fine-tuned to the specific intricacies of the task.

This approach for training has some benefits, as elucidated in Table I. The pretrained models yielded not only high accuracy, but also swiftness and compactness, making them highly favorable options for the present study's objectives. By building upon the foundations laid by these models, the study achieved a synthesis of existing knowledge and new insights, culminating in an efficient and effective pain recognition framework.

The process of transfer learning within this study was meticulously executed through the utilization of MATLAB's advanced Deep Learning Toolbox™. This specialized toolbox offers a robust framework that empowers researchers and practitioners to seamlessly design and deploy deep neural networks (DNNs), leveraging a repertoire of algorithms, pretrained networks, and model application methodologies. The symbiotic integration of these resources streamlines the transfer learning process and optimizes its implementation.

The study leveraged the prowess of pretrained networks as a cornerstone of the transfer learning journey. These pretrained networks served as foundational building blocks, upon which the study's specific pain recognition task was constructed. The agility and efficiency offered by these pretrained networks significantly expedite the learning process, allowing the model to swiftly adapt to the intricacies of the new dataset. Transfer learning is commonly used in deep learning applications. A pretrained network can be taken and used as a starting point for learning a new task. Perfecting a network with transfer learning is much faster and easier than practicing from scratch. The advantage of transfer learning is that the pretrained network has already learned a rich set of features that can be applied to a wide range of other similar tasks. Transfer learning offers several advantages: it enables to transfer the learned features of a pretrained network to a new problem, it is faster and easier than training a new network, it reduces training time and dataset size, and it perform deep learning without needing to learn how to create a whole new network [20].

To pave the way for the deployment of these pretrained models within the MATLAB environment, an essential preliminary step involves the installation of the respective pretrained model via the add-ons command [21]. This installation procedure is crucial for facilitating a seamless integration of the model into the MATLAB framework, thereby enabling its application in the subsequent phases of the study. The meticulous orchestration of this transfer learning process within the MATLAB's Deep Learning Toolbox™ underscores the study's commitment to leveraging cutting-edge tools and techniques for the attainment of its research objectives.

TABLE II
PARAMETERS USED IN THE DESIGN OF DENSENET201, MOBILENETV2, RESNET101 AND EFFICIENTNETB0 MODELS

Parameter	DenseNet201, MobileNetV2, ResNet101, and EfficientNetb0 Models
Epoch	100
Minibatch	24
Optimizer	Stochastic gradient descent (SGD)
Initial Learning Rate	10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}
Momentum Optimizer	0.9
WeightLearnRateFactor	10
BiasLearnRateFactor	10

In addition, fine-tuning was done by adjusting the hyperparameter values consisting of initial learning rate (ILR), maximum number of epochs, minibatch size, momentum, and optimizer, taking into account the minimum duration and maximum accuracy values. The parameters used in designing the DenseNet201, MobileNetV2, ResNet101, and EfficientNetb0 models are shown in Table II. The model was trained with epochs of 100, using a momentum of 0.9 with a minibatch size of 24. WeighLearnRateFactor and BiasLearnRateFactor parameters were set at 10, using the stochastic gradient descent (SGD) optimizer with an initial learning rate set of 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} [21]. Learning rate is one of the training parameters to calculate weight correction values during the training process [18]. The most ideal learning rate value is the value that produces the optimal level of accuracy and does not require a long training time.

D. DATA TESTING

The evaluation of data testing within this study adhered to the outcomes generated during the training process, leveraging the optimal deep learning models, and learning rates that exhibited the highest degree of accuracy. The assessment of testing results involved a comprehensive analysis, encompassing accuracy, recall, precision, and F1-score critical evaluation metrics computed through the following equation:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity\ (Recall) = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1-Score = 2 \times \frac{recall \times precision}{recall+precision} \quad (4)$$

Equations (1) through (4) are instrumental in the calculation of the accuracy value, a pivotal indicator of the model's prowess in categorizing images. Within this context, each equation encapsulates distinct aspects of the model's performance, providing a holistic view of its classification capabilities.

The true positive (TP) component signifies the accurate classification of a positive sample, denoting the successful identification of an image depicting pain. Conversely, the true negative (TN) component indicates the correct classification of a negative sample, symbolizing an image devoid of pain. The false positive (FP) scenario transpires when a negative sample is erroneously categorized as positive, while the false negative (FN) arises when a positive sample is mistakenly labeled as negative. These four elements illuminate the intricate interplay of classification outcomes.

TABLE III
COMPARISON OF DEEP LEARNING MODEL TO LEARNING RATE

		Deep Learning Model			
		DenseNet201	**EfficientNetb0**	**MobileNetV2**	**ResNet101**
Learning Rate of 10^{-1}	**VA (%)**	93.00	96.20	99.00	97.00
	ET	2,039 min and 28 s	558 min and 20 s	325 min and 32 s	2,219 min and 37 s
Learning Rate of 10^{-2}	**VA (%)**	99.60	99.60	99.60	99.00
	ET	2,000 min and 52 s	578 min and 25 s	312 min and 28 s	3,508 min and 2 s
Learning Rate of 10^{-3}	**VA (%)**	99.60	99.40	99.60	99.60
	ET	2,138 min and 18 s	579 min and 41 s	336 min and 27 s	3,501 min and 14 s
Learning Rate of 10^{-4}	**VA (%)**	99.40	99.20	99.20	99.00
	ET	2,132 min and 25 s	585 min and 44 s	342 min and 31 s	3,524 min and 27 s
Learning Rate of 10^{-5}	**VA (%)**	99.40	85.60	99.00	99.00
	ET	2,100 min and 42 s	558 min and 37 s	330 min and 13 s	3,529 min and 15 s

Notes:
VA = Validation accuracy (100%)
ET = Elapsed time

The overarching metric of accuracy encapsulates the system's comprehensive performance across all data points, underscoring its capacity to correctly categorize both positive and negative instances. Sensitivity (also known as true positive rate or recall), a parameter entailing the accurate classification of positive data, quantifies the system's adeptness at identifying individuals exhibiting positive attributes (in this case, experiencing pain).

Positive predictive value (PPV) is a statistical measure that indicates how likely it is that a positive test result is correct. It is calculated by dividing the number of TP by the total number of positive test results. PPV played a pivotal role in assessing the model's precision in identifying pain-related images within the realm of total positive class predictions. This metric provides insights into the system's accuracy in pinpointing instances of pain within the broader context of positive predictions [13], [22]. Through the integration of these equations and metrics, the study not only gauged the model's effectiveness but also fostered a comprehension of its capabilities and limitations in pain recognition tasks.

III. RESEARCH RESULTS

The ultimate outcome of deep learning endeavors is intricately influenced by a multitude of factors, and one such significant determinant is the learning rate. Within the training phase, the study harnessed a meticulously designed deep learning model, subjecting it to a rigorous analysis that hinged upon a range of learning rates 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. The training process was iterated across 100 epochs, encompassing numerous cycles of learning and adaptation.

During this training stage, the model was immersed in a process of gradual refinement through exposure to the available training data. Following the training stage, the model's efficacy was scrutinized by assessing its performance on test data. The assessment was conducted using validation accuracy parameters and elapsed time, both integral facets that offer valuable insights into the model's efficiency and effectiveness.

The validation methodology employed the cross-validation, a technique characterized by the orchestration of multiple training iterations to derive the optimal value. This multifaceted approach ensures that the model's performance is meticulously evaluated, culminating in the extraction of the most optimal validation accuracy value from this iterative process.

The primary objective of this learning rate comparison is to ascertain the most optimal configuration that yields exceptional performance outcomes, characterized by increased accuracy and efficient training times. Through a comprehensive analysis and evaluation of the interplay between learning rates, validation accuracy, and elapsed time, the study aims to unlock the best possible amalgamation, equipping researchers and practitioners to conduct effective deep learning endeavors in the domain of pain recognition.

The training outcomes stemming from the interplay of deep learning models and varying learning rates are comprehensively detailed in Table III. During the training phase, a learning rate of 10^{-2} emerged as particularly impactful, yielding remarkable validation accuracy results across DenseNet201, EfficientNetb0, and MobileNetV2 models. Specifically, these models exhibited an impressive accuracy of 99.60%, coupled with swift elapsed times of 2,000 min and 52 s for DenseNet201, 578 min and 25 s for EfficientNetb0, and 312 min and 28 s for MobileNetV2.

However, ResNet101 model demonstrated an exceptional performance, attaining the highest accuracy of 99.60%. This performance was achieved at a learning rate of 10^{-3}, an elapsed time of 3,501 min, and 14 s.

From these insights, it becomes evident that the optimal learning rate for the deep learning model, specifically MobileNetV2, is 10^{-2}. This meticulous analysis underscores the interplay between learning rates, model architectures, and performance metrics, ultimately unveiling the ideal configuration for the task of pain classification through facial expressions.

Furthermore, as the study transitions to the data testing phase, a new dataset distinct from the training data was employed. This stage stands as a crucial juncture for evaluating the pain classification model's performance via facial expressions. The evaluation process leveraged a confusion matrix to gauge the model's sensitivity, precision, and F1-score values.

To conduct this evaluation, the learning rate of 10^{-2} was uniformly employed across DenseNet201, EfficientNetb0, and

TABLE IV
CONFUSION MATRIX OF THE DEEP LEARNING MODELS

Model	Accuracy (%)	Sensitivity	Precision	F1-Score
DenseNet201	81.1	81.06	81.52	81.29
EfficienNetb0	78.2	78.24	78.46	78.35
MobileNetV2	79.4	79.38	80.04	79.80
ResNet101	78.2	78.24	78.58	78.41

MobileNetV2 models. In contrast, the ResNet101 model used a learning rate of 10^{-3}. This decision was done considering the model's ability to achieve the highest accuracy during the training phase. By meticulously fine-tuning the learning rate parameter, the study aims to optimize the model's full potential and achieve optimal performance outcomes in the domain of pain classification.

The outcomes of the confusion matrix analysis, as manifested in Table IV, provide an essential insight on the performance of the DenseNet201, EfficientNetb0, MobileNetV2, and ResNet101 models. These models, each representing distinct architectural approaches, were subjected to rigorous testing with the dataset, yielding insights into their classification capabilities.

The results in Table IV demonstrates that the DenseNet201 architecture achieved the highest accuracy of 81.1% among the models analyzed, making it the top-performing model. Moreover, it exhibited remarkable sensitivity, precision, and F1-score values of 81.06, 81.52, and 81.29, respectively. This comprehensive display of high metrics underscores the DenseNet201's efficacy in pain classification through facial expressions.

Although DenseNet201 took the lead, the MobileNetV2, EfficientNetb0, and ResNet101 models also contributed significantly to the domain. Their respective accuracies of 78.2%, 79.4%, and 78.2% highlight their competence in the task of pain classification. These models, while not surpassing the DenseNet201 architecture, remain competitive with performances that are notably close to each other.

The nuanced insights gleaned from these results not only corroborate the effectiveness of various architectural choices but also provide a foundation for making informed decisions about model selection based on specific use cases and requirements. The comprehensive evaluation of these architectures elucidates the strengths and areas for improvement of each model, ultimately guiding the broader landscape of pain classification research.

IV. CONCLUSION

The study presents a comprehensive exploration of pain recognition methodologies, successfully combining the Viola-Jones technique and an array of four advanced deep learning models, namely MobileNetV2, EfficientNetb0, ResNet101, and DenseNet201. The process encompasses a meticulous calibration of learning rate values spanning from 10^{-1} to 10^{-5}.

A prominent pattern was attained from analyzing the training data. A learning rate of 10^{-2} served as a catalyst for heightened accuracy validation across the DenseNet201, EfficientNetb0, and MobileNetV2 models, each attaining an impressive accuracy level of 99.60%. Importantly, these models achieved this remarkable performance without compromising on elapsed time, with processing times of 2,000 min and 52 s, 578 min and 25 s, and 312 min and 28 s, respectively.

Concurrently, the ResNet101 yielded the highest accuracy of 99.60%, achieved at a learning rate of 10^{-3}. This achievement, however, entailed an extended elapsed time of 3,501 min and 14 s, showcasing a trade-off between accuracy and processing efficiency.

Subsequent examination of the test results on a different dataset sheds light on the practical implications of the models' performance. Within this evaluation framework, the DenseNet201 model demonstrated superior performance, securing an accuracy of 81.1%. The results of the confusion matrix accentuated this achievement, with impressive sensitivity, precision, and F1-score values of 81.06, 81.52, and 81.29, respectively. Notably, the MobileNetV2, EfficientNetb0, and ResNet101 models also contributed significantly, yielding accuracies of 78.2%, 79.4%, and 78.2%, respectively.

In synthesis, this multifaceted investigation underscores the paramount importance of selecting an optimal learning rate for deep learning models. The findings underscore the efficiency of a learning rate of 10^{-2} when coupled with the MobileNetV2 model, yielding a harmonious blend of accuracy and operational efficiency. By harnessing a strategic amalgamation of methodologies and models, the study propels the field of pain recognition through facial expressions, carving a path toward increasingly effective and efficient methodologies.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest in the research and preparation of this paper.

AUTHORS' CONTRIBUTIONS

Conceptualization, Raihan Islamadina, Khairun Saddami, Fitri Arnia, Taufik Fuadi Abidin, and Rusdha Muharar; methodology, Raihan Islamadina, Khairun Saddami, Fitri Arnia, Taufik Fuadi Abidin, and Rusdha Muharar; software, Raihan Islamadina, Khairun Saddami, and Muhammad Irwandi; writing—original draft preparation, Raihan Islamadina; writing—reviewing and editing, Raihan Islamadina, Khairun Saddami, Fitri Arnia, Taufik Fuadi Abidin, Rusdha Muharar, and Aulia Syarif Aziz.

ACKNOWLEDGMENT

This research was funded by DIPA UIN AR-Raniry Banda Aceh in 2023. Thank you to the advisors from Syiah Kuala University in Banda Aceh and the research team from UIN Ar-Raniry Banda Aceh who had contributed to conducting this research and writing the article.

REFERENCES

[1] K.D. Craig, "The facial expression of pain better than a thousand words?" *APS J.*, vol. 1, no. 3, pp. 153–162, 1992, doi: 10.1016/1058-9139(92)90001-S.

[2] M.A. Lazarini, R. Rossi, and K. Hirama, "A systematic literature review on the accuracy of face recognition algorithms," *EAI Endorsed Trans. IoT*, vol. 8, no. 30, pp. 1–11, Sep. 2022, doi: 10.4108/eetiot.v8i30.2346.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2001, pp. I-511–I-518, doi: 10.1109/CVPR.2001.990517.

[4] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.

[5] M.F. Hirzi, S. Efendi, and R.W. Sembiring, "Literature study of face recognition using the Viola-Jones algorithm," *2021 Int. Conf. Artif. Intell. Mechatronics Syst. (AIMS)*, 2021, pp. 1–6, doi: 10.1109/AIMS52415.2021.9466010.

[6] F. Elgendy, M. Alshewimy, and A. Sarhan, "Pain detection/classification framework including face recognition based on the analysis of facial expressions for e-health systems," *Int. Arab J. Inf. Technol.*, vol. 18, no. 1, pp. 125–132, Jan. 2021, doi: 10.34028/iajit/18/1/14.

[7] R.M. Al-Eidan, H. Al-Khalifa, and A. Al-Salman, "Deep learning-based models for pain recognition: A systematic review," *Appl. Sci.*, vol. 10, no. 17, pp. 1–15, Aug. 2020, doi: 10.3390/app10175984.

[8] M.N. Chaudhari, M. Deshmukh, G. Ramrakhiani, and R. Parvatikar, "Face detection using Viola Jones algorithm and neural networks," *2018 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697768.

[9] P. Rodriguez *et al*., "Deep pain: Exploiting long short-term memory networks for facial expression classification," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3314–3324, May 2022, doi: 10.1109/TCYB.2017.2662199.

[10] D.L. Martinez, O. Rudovic, and R. Picard, "Personalized automatic estimation of self-reported pain intensity from facial expressions," *2017 IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017, pp. 2318–2327, doi: 10.1109/CVPRW.2017.286.

[11] G. Bargshady *et al*., "Ensemble neural network approach detecting pain intensity from facial expressions," *Artif. Intell. Med.*, vol. 109, pp. 1–12, Sep. 2020, doi: 10.1016/j.artmed.2020.101954.

[12] G. Bargshady *et al*., "The modeling of human facial pain intensity based on temporal convolutional networks trained with video frames in HSV color space," *Appl. Soft Comput.*, vol. 97, pp. 1–14, Dec. 2020, doi: 10.1016/j.asoc.2020.106805.

[13] R. Islamadina *et al*., "Performance of deep learning benchmark models on thermal imagery of pain through facial expressions," *2022 IEEE Int. Conf. Commun. Netw. Satell. (COMNETSAT)*, 2022, pp. 374–379, doi: 10.1109/COMNETSAT56033.2022.9994546.

[14] G. Huang, Z. Liu, L.V.D. Maaten, and K.Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[15] M. Sandler *et al*., "MobileNetV2: Inverted residuals and linear bottlenecks," *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[17] M. Tan and Q.V. Le, "EfficientNet: Rethinking model scaling for convolution neural networks," 2019, *arXiv.1905.11946*.

[18] J. Brownlee (2020) "Understand the Impact of Learning Rate on Neural Network Performance," [Online], https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/, access date: 22-Jun-2023.

[19] M.A. Haque *et al*., "Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities," *2018 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG 2018)*, 2018, pp. 250–257, doi: 10.1109/FG.2018.00044.

[20] M.H. Beale, M.T. Hagan, and H.B. Demuth, *Deep Learning Toolbox™ User's Guide*. (2020). Access date: 19-Aug-2023. [Online]. Available: https://www.mathworks.com/help/deeplearning/index.html

[21] A. Schoenauer-Sebag, M. Schoenauer, and M. Sebag, "Stochastic gradient descent: Going as fast as possible but not faster," 2017, *arXiv:1709.01427*.

[22] A.G. Lalkhen and A. McCluskey, "Clinical tests: Sensitivity and specificity," *Contin. Educ. Anaesth. Crit. Care Pain*, vol. 8, no. 6, pp. 221–223, Dec. 2008, doi: 10.1093/bjaceaccp/mkn041.