Ensemble Learning-Based Pain Intensity Identification Model Using Facial Expressions

Abdul Rahaman Wahab Sait1,* and Ashit Kumar Dutta2

1Department of Documents and Archive, Center of Documents and Administrative Communication, King Faisal University, Hofuf 31982, Al-Ahsa, Saudi Arabia
2Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Ad Diriyah, Riyadh 13713, Saudi Arabia

Correspondence to:
Abdul Rahaman Wahab Sait*, E-mail: asait@kfup.edu.sa, Mobile: +966543377539

Received: January 22 2024; Revised: March 13 2024; Accepted: March 13 2024; Published Online: March 28 2024

ABSTRACT

The pain identification model (PIM) can assist the healthcare professionals to render effective services to individuals. Disabled individuals can benefit from the automated PIM. Ensemble learning is widely employed for developing medical applications. A model for classifying the pain intensity using facial expression images is proposed in this study. A ShuffleNet V2 model is fine-tuned to extract features using fusion feature and class activation map techniques. CatBoost and XGBoost models are used as base models to predict pain intensities. The authors used the support vector machine (SVM) model as a meta-model to produce a final outcome. They optimize the SVM model in order to identify pain using the predictions of the base models. The model is generalized using the University of Northern British Columbia–McMaster dataset. The dataset encompasses 200 videos and 48,000 annotated images. The comparative analysis outcome highlights the exceptional performance of the proposed PIM. An optimal accuracy of 98.7% and an F1-score of 98.0% indicate the effectiveness of the proposed model. The uncertainty analysis outcome revealed that the model is reliable and can be deployed in healthcare centers. However, substantial training is required to boost the efficiency of the proposed model in real-time settings.

KEYWORDS

pain identification, ensemble learning, ShuffleNet V2, CatBoost, XGBoost, facial expression, hyperparameter tuning

INTRODUCTION

Pain is a vital neural system signal that demands medical attention (Gkikas and Tsiknakis, 2023). As a consequence, the patient has discomfort due to the perplexing and ambiguous sensation. An individual’s sensitivity to pain may vary considerably depending on their age, gender, and the specific damage (Salekin et al., 2021). Pain assessment and management are critical for many medical disorders and treatments. Pain is frequently assessed subjectively by patients, caregivers, or medical staff. Despite their utility and convenience, subjective reports have limits. Patients in the pediatric range, those with neurological impairments, dementia, post-traumatic stress disorder, or who need mechanical ventilation are unable to self-report (Alghamdi, 2023). Self-reported pain assessment may be difficult to understand and demands expert observers to quantify manually (Ismail and Waseem, 2023). However, it is limited by measurement inconsistency across scale dimensions, sensitivity to suggestion, impression management or deception, and physicians’ and patients’ pain concept analysis. Pain identification models (PIMs) typically employ scales, assessment tools, or datasets to measure pain intensity, duration, and additional characteristics (Semwal and Londhe, 2021). These models may predict or diagnose pain using physiological signs, facial expressions, speech, or other modalities. Biomedical studies have shown that facial expressions precisely convey levels of discomfort. Machine learning methods are being used to build automatic nonverbal pain-detecting systems (Semwal and Londhe, 2021).

Previous investigations have demonstrated that facial action units (AUs) may be detected for autonomous pain expression evaluation (Alghamdi and Alaghband, 2023). On the contrary, existing methods for identifying facial pain AUs are restricted to regulated circumstances. The inability to properly measure pain in severely sick patients is the primary barrier to providing them with suitable therapy. To regulate opioid dosages and the healing process, pain evaluation...
is essential in the intensive care unit (ICU) (Huang et al., 2023). The characteristics of traumatic events vary by individual, making pain assessment complicated. Individual patient self-reporting is the gold standard for pain evaluation in the ICU. Despite their prevalence, popular self-report pain instruments such as the visual analog scale and the numeric rating scale are subjective and susceptible to bias (Huang et al., 2023).

In addition to facial expressions, bodily movements, gestures, and audio can be used for pain assessment (Yuan et al., 2023). Images or videos of face and body expressions may be recorded using various optical and light sensors (Yuan et al., 2023). In order to provide additional visual channels, the majority of researchers utilize the capabilities of color RGB cameras (Sri Shreya et al., 2023). However, in some instances, depth and thermal camera sensors are used. Many physiological modalities use biosignals to record electrical activity from tissues and organs (Sri Shreya et al., 2023). The PIM plays a significant role in rehabilitation facilities to improve the quality of life of disabled individuals (De Sario et al., 2023). Pain may be challenging to communicate for disabled individuals, especially those with speech limitations. When verbal communication is restricted, facial expressions and behavioral clues are essential for pain intensity recognition. Improving rehabilitation results requires individualized treatment (Gouverneur et al., 2023). A precise pain intensity evaluation allows healthcare and rehabilitation experts to customize strategies for each impaired patient (Gouverneur et al., 2023). Improving methods for reducing pain for people with disabilities depends on timely and accurate pain assessments (Lima et al., 2023). Pain intensity identification and management minimize subsequent problems and guarantee safe rehabilitation (Delgado et al., 2018).

The facial anatomy-based Facial Action Coding System (FACS) processes facial AUs to capture immediate facial expression variations (Delgado et al., 2018). Facial expression may be expressed by a combination of facial AUs, which are defined by the relative contraction or relaxation of facial muscles. Based on the FACS, Prkachin and Solomon established the Prkachin Solomon Pain Index (PSPI) score (Delgado et al., 2018). The PSPI score accounts for the existence and severity of action units AU-4 (brow lowerer), AU-6 (cheek raiser), AU-7 (lid tightener), AU-9 (nose wrinkler), AU-10 (upper lip raiser), and AU-43 (eyes closed) on a scale of 0-5 (Egede et al., 2020). A pain scale with 16 points is generated by adding the existence of AUs to their respective intensities. This approach requires skilled individuals to manually code AUs, which is time-consuming and expensive. However, the existing datasets with PSPI can be used to develop deep learning (DL) models to detect pain intensities using facial images.

Clinical workflow may benefit from an autonomous, real-time pain detection system. An autonomous pain expression evaluation system requires AU identification (Chen et al., 2018; Al-Eidan et al., 2020; Hassan et al., 2021; Werner et al., 2022). The lack of an annotated dataset in uncontrolled situations limits autonomous pain detection system development. Video recordings were taken in the past by researchers from McMaster University and the University of Northern British Columbia (UNBC) of the faces of 25 volunteers who were experiencing shoulder discomfort with a variety of diagnoses (Lucey et al., 2011). Certified FACS coders coded each video frame in a controlled environment (Mavadati et al., 2013). Recent datasets, such as BP4D–Spontaneous and BP4D+, have healthy participants’ facial AUs annotated and pain triggered by the cold pressor task. The existing pain expression datasets are controlled or semi-controlled (Mavadati et al., 2013; Menchetti et al., 2019).

Transfer learning and information from varied sources are achievable using pre-trained models (Menchetti et al., 2019). It becomes exceedingly valuable in training automated PIMs with limited labeled data. Ensemble models typically outperform individual models (Bargshady et al., 2020). An ensemble of models with different architectures or training procedures may capture additional patterns (Semwal and Londhe, 2021). Ensemble learning (EL) makes more resilient and generalizes to unique patterns. When individual models are susceptible to detecting noise or certain patterns in the training data, ensemble approaches may assist in reducing overfitting (Xu and de Sa, 2020; El Morabit et al., 2021; Susam et al., 2021). The diversity of ensemble models mitigates the impact of individual model biases, resulting in improved performance on real-time data.

Identifying pain is a crucial responsibility in healthcare since precise and prompt evaluation of pain is necessary for successful pain control and therapy. Conventional pain evaluation approaches often depend on subjective self-reporting or clinical observations, which may be susceptible to biases and mistakes. Moreover, current PIMs may encounter difficulties in applying their findings to a wide range of patient demographics and clinical environments. Robust and generalizable PIMs are necessary to accurately and consistently identify and measure pain in diverse persons and situations.

The rationale for creating a pain detection model using EL stems from the constraints of current methods and the possible advantages of ensemble techniques in overcoming these difficulties. EL offers a promising avenue for improving PIMs’ accuracy, robustness, and generalization by leveraging the diversity among multiple base models. EL, by combining predictions from many models, may effectively address overfitting, decrease model variation, and improve the dependability of pain diagnosis across multiple patient demographics, clinical situations, and assessment contexts. Moreover, EL facilitates incorporating supplementary data from many sources, including physiological signals, facial expressions, and behavioral cues, resulting in a more thorough and precise pain evaluation. An EL-based PIM can potentially optimize clinical decision-making, improve patient outcomes, and extend our comprehension of pain assessment in healthcare settings.

There have been promising developments in computer vision, machine learning, and signal processing that could potentially lead to complex models to draw valuable insights from unstructured data. The dependability and precision of automated pain detection may be improved using these technologies. Cultural sensitivity, privacy, and autonomy must be considered while designing the model to address pain assessment biases and ethical issues. The automated PIM should
be tested effectively in clinical settings. Generalizing the model in various pain circumstances and integrating it into present workflows are required to develop an effective PIM. These factors motivated the authors to develop an EL-based PIM using facial images. The study contributions are:

- A fine-tuned ShuffleNet V2 model-based feature extraction model using facial expression images.
- An EL approach using CatBoost, XGBoost, and random forest models for identifying pain levels.

The authors have conducted a systematic literature review in order to identify the existing PIM. Chen et al. (2018) emphasized the importance of DL techniques and FACS in assessing pain. However, they reported a limited number of DL techniques. Hassan et al. (2021) discussed the application of DL-based PIMs in healthcare centers. Werner et al. (2022) presented the effectiveness of the existing automated PIMs. Al-Eidan et al. (2020) reported the features and limitations of automated pain assessment models. Pain may be associated with potential tissue damage (Al-Eidan et al., 2020). It can be characterized based on severity, duration, and distribution. Automated pain assessment tools demand necessary datasets in order to identify pain without any complexities (Lucey et al., 2011). Training a DL method requires substantial computational resources to detect pain using facial expression images. PIM implementation in clinical settings causes challenges to the healthcare professionals. Thus, the PIM developers need to maintain a trade-off between computational power and the model’s prediction.

A limited number of datasets may reduce the PIM’s performance. The UNBC-McMaster shoulder pain dataset, the BioVid heat pain dataset, and the Denver Intensity of Spontaneous Facial Action dataset are widely applied for building PIM (Mavadati et al., 2013). Moreover, these datasets are freely available for the researchers to extend their research in pain assessment. Training and inference are the two stages in the operation of DL-based systems. Graphics processing units (GPUs) are required during the training phase due to the high computing demands of the entire process (Mavadati et al., 2013). Inference uses the learned model to predict novel information. The central processing unit usually handles this procedure; however, numerous variables determine the hardware (Menchetti et al., 2019). For instance, real-time applications with crucial latency demand greater requirements compared to offline approaches with later predictions. In addition, the amount of operations and the trained model’s properties, such as floating-point operations per second (FLOPS), should be considered for evaluating the efficiency of PIM (Bargshady et al., 2020).

Menchetti et al. (2019) proposed a two-stage DL model for detecting pain. An end-to-end DL-based automated facial expression recognition model was developed. The video clips were processed frame by frame to generate AU likelihood values. The study’s findings indicated that the model has outperformed the state-of-the-art PIM.

Bargshady et al. (2020) recommended a PIM to identify pain levels. They followed the EL approach for predicting pain levels using the facial images. They employed VGGFace and principal component analysis to make the prediction. The generalization of the model was performed using the UNBC–McMaster dataset. The model achieved an average accuracy of 89% and a receiver operating characteristic of 93%.

Semwal and Londhe (2021) introduced a convolutional neural network (CNN)-based model for detecting pain intensities. They used VGGNet, MobileNet, and GoogleNet models as base models for classifying the images. The average ensemble rule was used to integrate the CNN results to make a final prediction. The model generated results with an F1-score of 91.41%.

El Morabit et al. (2021) compared the effectiveness of the CNN models in identifying the pain levels using facial expressions. They used MobileNet, GoogleNet, ResNext 50, ResNet 18, and DenseNet 161 models for image classification. The study’s outcome indicated the usefulness of the automated PIM.

Xu and de Sa (2020) introduced a PIM using the EL approach. A multi-task learning neural network was used to generate a pain score. The pain scores were linearly combined using the EL model to produce a final score. The model achieved an average error of 1.73. Similarly, Susam et al. (2021) presented an approach integrating subjective self-reporting and video facial expressions to classify pain intensities. They trained the model using the videos of children who recovered from a laparoscopic appendectomy. The findings revealed that the integrated approach yielded better results than the individual approach.

Psychosocial variables, including nervousness, depression, and stress, influence pain. A more comprehensive approach and a more excellent knowledge of pain and mental health are essential to include these factors in automated evaluations. Pain is a complex and subjective sensation that is affected by a range of elements, including individual variations, cultural influences, and environmental signals. In order to be clinically useful, DL models must be able to comprehend and encompass the intricacy of various forms of pain, such as acute versus chronic, nociceptive versus neuropathic, and be applicable to varied populations. DL models used for pain identification often lack interpretability, which poses a challenge in comprehending the specific factors that influence the model’s predictions and in having confidence in their clinical reliability. Research is required to construct interpretable DL models that may provide insights into the pain detection process and aid in therapeutic decision-making. The absence of objective pain metrics renders automated approaches extremely difficult to validate.
MATERIALS AND METHODS

The authors build a PIM using an ensemble approach. Figure 1 illustrates the methodology of the proposed PIM. The facial images are utilized to train the model. A feature extraction technique based on the ShuffleNet V2 model is suggested to extract the intricate patterns. Ensemble models use the variations across individual models to reflect distinct characteristics of the underlying data distribution. Individual base models may demonstrate expertise in recognizing specific patterns or correlations in the data. EL enables the integration of these complementing patterns, resulting in enhanced generalization performance. It reduces overfitting by combining predictions from various base models, which may have been trained on different subsets of the data or using different regularization approaches. Ensemble models may provide more reliable and adaptable predictions by merging forecasts from numerous models, reducing vulnerability to overfitting. The authors used CatBoost and XGBoost as base models to generate an outcome using the extracted features. Finally, they employ the support vector machine (SVM) model as a meta-model to classify the facial expression into multiple classes including PSPI = 0, PSPI = 1, 2 ≤ PSPI ≤ 3, and PSPI > 3. The lightweight architecture of the ShuffleNet V2 model enables the implementation of feature extraction with limited resources. The existing studies highlighted the significance of CatBoost, XGBoost, and SVM models in image classification. These factors motivated the authors to apply these models in the proposed study.

Image acquisition

To train the proposed model, the authors employ the UNBC–McMaster dataset. The dataset provides videos of individuals who were suffering from shoulder pain. The videos were recorded while individuals performed a series of active and passive range of motion tests on their limbs in two different scenarios. The certified FACS coders were employed to code each frame. The dataset contains 200 video sequences of spontaneous facial expressions, 48,398 FACS-coded frames, sequence-level self-reports, and observer measures.

Feature extraction

The authors construct a CNN model with five convolution layers and batch normalization. They train the model using the ShuffleNet V2 model’s weights. Figure 2 shows the proposed feature extraction process. The focus on processing efficiency and parameter reduction distinguishes ShuffleNet V2 from other CNN models. The ability to shuffle channels is a unique feature of ShuffleNet V2. The expressive capability of the network is increased by promoting information interchange between layer channels. In order to extract features, ShuffleNet V2 employs a stack of convolutional layers. The convolutional layers recognize low-level information in input images using filters. Convolutional procedures capture...
the local patterns. Each input channel group is convolved separately in group convolution. This simplifies the process of computation and lowers the amount of parameters. The authors used element-wise sum or average feature fusion to add or average feature map components. Information from multiple levels is combined using this lightweight method. In addition, the authors integrate a class activation map (CAM) to visualize the crucial patterns. Subsequently, a fully connected layer and a reshape function are included with the feature extraction model for generating a two-dimensional (2D) vector.

In this study, the authors follow a stacking generalization technique that combines the predictions from multiple base models to increase predictive performance. They employ SVM as a meta-model to generate outcomes by incorporating base model predictions. They train the CatBoost (https://github.com/catboost/catboost) and XGBoost (https://github.com/topics/xgboost) models in order to obtain a prediction. The predictions are used as a feature to train the SVM classifier. The SVM classifier learns to identify the crucial patterns of the features and classify them into multiple classes. The proposed PIM is presented in Figure 3.

Unlike other gradient-boosting methods, CatBoost requires no preprocessing, encoding, or one-hot encoding to handle categorical information. As a result, the categorical data can be processed with limited computational resources. CatBoost provides powerful GPU acceleration and parallelization, rendering it ideal for massive datasets and complex models. The approach uses ordered boosting and random permutations in order to prevent overfitting and improve generalization. Without imputation, CatBoost handles missing data. During training, oblivious trees handle missing values. With its speed and efficiency, CatBoost can train faster than other gradient-boosting frameworks. However, the authors apply the Hyperband optimization technique to fine-tune the performance of the CatBoost model in identifying pain.

Regularization procedures, including L1 (LASSO) and L2 (ridge) regularization, are included in the objective function of XGBoost. This reduces overfitting and increases model generalization. Because of its architecture, XGBoost can expand over several processors or nodes and effectively analyze facial images. Developers may construct custom objective functions in XGBoost for specific tasks or challenges. This feature allows the authors to improve the efficiency of the XGBoost model using the Hyperband optimization. XGBoost was developed to manage missing data during the training phase. Based on data, it can automatically handle missing values. To improve model generalization, the authors regularize tree building to manage tree complexity. XGBoost’s built-in cross-validation assists in tracking the model performance during training and minimizes overfitting.

In the context of PIM, SVM processes a 2D vector to make a final decision. The authors employ the radial basis function in order to process nonlinear kernels. For instance, let $f_1$ and $f_2$ be the features, and $C$ be the class. Equation 1 shows the pain level classification.

$$C = SVM(f_1, f_2, K, R, G)$$

where SVM is the classifier, $K$ is the kernel function, $R$ is the regularization parameter, and $G$ is the gamma value.

The objective of the SVM classifier is to identify a hyperplane that classifies the data points of pain levels. Equation 2 presents the hyperplane computation associated with each image classification.

$$w_1 f_1 + w_2 f_2 + b = 0$$

where $w_1$ and $w_2$ are the weights associated with features $f_1$ and $f_2$, and $b$ is the bias term.

Equation 3 highlights the mathematical form of the decision function of PIM.

$$F(f) = w f + b$$

where $w$ is the vector’s weight, $f$ is the feature vector, and $b$ is the bias term.

The authors apply the kernel tricks to operate the classifier in a higher dimensional space. This process supports the proposed model in handling nonlinearly separable data. The authors apply the fivefold cross-validation with unique hyperparameter settings to fine-tune the SVM classifier. This process prevents overfitting and objectively estimates the model’s performance. In addition, the authors use the Min-Max scaling technique to normalize the features.

The benchmark metrics, including accuracy (Acc.), precision (Pre.), recall (Rec.), and F1-score, are used for performance evaluation. A model’s accuracy may be defined as the degree to which its predictions are accurate. It is the

![Figure 3: The suggested fine-tuned SVM classifier. Abbreviation: SVM, support vector machine.](image-url)
ratio of accurately predicted instances to total instances. In a balanced context, accuracy shows the model’s performance. Although accuracy is a valuable indicator for evaluating the performance of a model at a high level, it is essential to supplement it with additional measures, particularly in situations where there is an imbalance across classes or where various kinds of mistakes have diverse consequences. For pain detection models, precision is paramount, especially in healthcare settings, where the consequences of incorrect positives could be life-threatening.

As part of a holistic assessment method, it shows the model’s capacity to produce accurate positive predictions. Precision has a positive predictive value. The model’s dependability is measured when it predicts a positive class instance. A high precision means the model makes reliable positive predictions. Recall or true-positive rate assesses its ability to detect positive events among all real positive instances. It shows the ability of a PIM to capture all pain episodes. A high recall implies that the model minimizes false negatives, ensuring actual pain is not missed. The harmonic mean of accuracy and recall is F1-score. It provides a balanced model performance measure, including false positives and negatives. False positives and negatives affect pain identification. False positives may lead to unwarranted therapies, while false negatives may overlook pain. These errors are balanced by recall and F1-score. Patient outcomes could be directly affected by decisions made in the medical field based on model predictions. A balance between accuracy, recall, and F1-score ensures the trustworthiness of model prediction. In addition, the authors employ kappa (κ) to measure the model’s efficiency in imbalanced class settings. They compute computational strategies and perform uncertainty analysis to understand the model’s reliability.

RESULTS AND DISCUSSION

The experimental analysis is conducted using Windows 10 Professional, i7 5th generation, 16 GB RAM, NVIDIA R350X settings. The authors used fivefold cross-validation to train the proposed model. They employed PyTorch, TensorFlow, and Keras libraries for developing the PIM. They trained the model with 15 batches and 12 epochs. In addition, they extended the training up to 17 batches and 14 epochs. However, there was no significant improvement in the model’s performance. The authors fine-tuned the model using CAM and early stopping strategies. Table 1 displays the findings of fivefold cross-validation. The model has achieved a better outcome in each fold. It learned the crucial pain patterns and classified pain levels with high accuracy.

The suggested feature extraction has generated the key features to support the proposed model in making a decision. The diversity of the base models assisted the proposed model in achieving optimal accuracy.

The findings of the performance analysis of the proposed PIM are presented in Table 2. The multi-class classification ability is evaluated for individual classes. The model has obtained an average accuracy of 98.7% and an F1-score of 98.0%. The findings revealed that the suggested EL approach has produced an exceptional outcome. Figure 4 shows the performance of the proposed model’s multi-class classification ability. In addition, Figure 5 highlights the outcomes of the comparative analysis.

The current PIMs are compared with the proposed model. The proposed PIM outperforms the recent models by obtaining a superior outcome. The hyperparameter optimization has fine-tuned the proposed model to deliver an optimal result. In addition, the base models have produced a diverse outcome, which supports the meta-model to learn the intricate patterns of pain intensities. The outcomes of the comparative analysis are provided in Table 3.

Finally, the findings of the uncertainty analysis and computational strategies are listed in Table 4. The outcomes highlight the significance of the EL-based PIM in identifying the pain levels. Moreover, the proposed model required less number of parameters and floating-point operations (FLOPS) compared to the existing models.

The proposed EL-based PIMs enhance pain assessment accuracy and reliability, enabling healthcare practitioners to manage patients. It uses predictions from two base models, which employ effective techniques or data subsets. In comparison to individual models, EL models result in an improvement in the overall accuracy and resilience of the model. A patient’s pain level may be detected more accurately. The problem of overfitting, which is frequently encountered in DL, may be mitigated with the use of the proposed model. Ensemble approaches reduce training data noise and outliers by combining model predictions. This helps pain detection models generalize to new patient data. By using EL, numerous factors of pain representation may be captured by models. This improves model generalization to varied patient groups, ensuring it works effectively across demographics and diseases. The study’s outcome indicated that the proposed model could standardize pain assessment. Medical professionals may benefit from having access to an objective pain assessment instrument. The proposed model can assist in patient-centered treatment by offering reliable pain evaluations. With a better knowledge of the patient’s pain, physicians may personalize actions to enhance patient health.

Table 1: Fivefold cross-validation analysis outcome.

<table>
<thead>
<tr>
<th>Folds</th>
<th>Acc&lt;sub&gt;c&lt;/sub&gt;</th>
<th>Pre&lt;sub&gt;n&lt;/sub&gt;</th>
<th>Rec&lt;sub&gt;c&lt;/sub&gt;</th>
<th>F1-score</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95.1</td>
<td>94.2</td>
<td>94.1</td>
<td>94.1</td>
<td>89.4</td>
</tr>
<tr>
<td>2</td>
<td>96.7</td>
<td>95.7</td>
<td>95.2</td>
<td>95.4</td>
<td>89.7</td>
</tr>
<tr>
<td>3</td>
<td>96.9</td>
<td>96.8</td>
<td>96.5</td>
<td>96.6</td>
<td>90.5</td>
</tr>
<tr>
<td>4</td>
<td>97.4</td>
<td>97.8</td>
<td>98.1</td>
<td>97.9</td>
<td>91.3</td>
</tr>
<tr>
<td>5</td>
<td>98.7</td>
<td>98.0</td>
<td>97.9</td>
<td>98.0</td>
<td>93.5</td>
</tr>
</tbody>
</table>

Table 2: Multi-class performance analysis outcome.

<table>
<thead>
<tr>
<th>Folds</th>
<th>Acc&lt;sub&gt;c&lt;/sub&gt;</th>
<th>Pre&lt;sub&gt;n&lt;/sub&gt;</th>
<th>Rec&lt;sub&gt;c&lt;/sub&gt;</th>
<th>F1-score</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSPI = 0</td>
<td>98.5</td>
<td>98.1</td>
<td>98.3</td>
<td>98.2</td>
<td>93.7</td>
</tr>
<tr>
<td>PSPI = 1</td>
<td>98.8</td>
<td>97.9</td>
<td>97.6</td>
<td>97.7</td>
<td>92.9</td>
</tr>
<tr>
<td>2 ≤ PSPI ≤ 3</td>
<td>98.9</td>
<td>97.8</td>
<td>97.8</td>
<td>97.8</td>
<td>94.1</td>
</tr>
<tr>
<td>PSPI &gt; 3</td>
<td>98.7</td>
<td>98.5</td>
<td>98.2</td>
<td>98.3</td>
<td>93.5</td>
</tr>
<tr>
<td>Average</td>
<td>98.7</td>
<td>98.0</td>
<td>97.9</td>
<td>98.0</td>
<td>93.5</td>
</tr>
</tbody>
</table>

Abbreviation: PSPI, Prkachin Solomon Pain Index.
Figure 4: Multi-class performance analysis. Abbreviation: PSPI, Prkachin Solomon Pain Index.

Figure 5: Comparative analysis.

Table 3: Comparative analysis outcome.

<table>
<thead>
<tr>
<th>Model</th>
<th>Acc₀</th>
<th>Pre₀</th>
<th>Rec₀</th>
<th>F1-score</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susam et al. (2021)</td>
<td>97.1</td>
<td>95.8</td>
<td>95.7</td>
<td>95.7</td>
<td>90.1</td>
</tr>
<tr>
<td>Bargshady et al. (2020)</td>
<td>95.6</td>
<td>96.1</td>
<td>95.8</td>
<td>95.9</td>
<td>89.7</td>
</tr>
<tr>
<td>El Morabit et al. (2021)</td>
<td>94.1</td>
<td>94.5</td>
<td>94.3</td>
<td>94.4</td>
<td>90.9</td>
</tr>
<tr>
<td>Semwal and Londhe (2021)</td>
<td>94.8</td>
<td>94.6</td>
<td>94.5</td>
<td>94.5</td>
<td>91.2</td>
</tr>
<tr>
<td>Proposed PIM</td>
<td>98.7</td>
<td>98.0</td>
<td>97.9</td>
<td>98.0</td>
<td>93.5</td>
</tr>
</tbody>
</table>

Abbreviation: PIM, pain identification model.

Table 4: Uncertainty analysis and computational strategies.

<table>
<thead>
<tr>
<th>Model</th>
<th>Loss</th>
<th>CI</th>
<th>SD</th>
<th>Learning rate</th>
<th>Parameters (in millions)</th>
<th>FLOPS (in giga)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susam et al. (2021)</td>
<td>1.73</td>
<td>96.5-97.4</td>
<td>0.0004</td>
<td>$1 \times 10^{-4}$</td>
<td>27</td>
<td>42</td>
</tr>
<tr>
<td>Bargshady et al. (2020)</td>
<td>1.54</td>
<td>95.8-96.7</td>
<td>0.0003</td>
<td>$1 \times 10^{-3}$</td>
<td>34</td>
<td>56</td>
</tr>
<tr>
<td>El Morabit et al. (2021)</td>
<td>1.36</td>
<td>97.3-98.3</td>
<td>0.0005</td>
<td>$1 \times 10^{-3}$</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Semwal and Londhe (2021)</td>
<td>1.52</td>
<td>96.7-98.1</td>
<td>0.0004</td>
<td>$1 \times 10^{-2}$</td>
<td>26</td>
<td>42</td>
</tr>
<tr>
<td>Proposed PIM</td>
<td>1.15</td>
<td>98.4-98.9</td>
<td>0.0002</td>
<td>$1 \times 10^{-4}$</td>
<td>18</td>
<td>29</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; FLOPS, floating-point operations; PIM, pain identification model; SD, standard deviation.
satisfaction and care. As part of a decision support system, the proposed model may help healthcare providers determine pain management, medicine, and treatment plans.

CONCLUSION

A PIM was developed using the EL approach. The facial expression images were used to train the proposed model. The authors have improved the performance of the ShuffleNet V2 model-based feature extraction by integrating CAM and fusion feature techniques. The CatBoost and XGBoost models were used as base models to make predictions using the image features. The Hyperband optimization was employed to fine-tune the base models. The authors employed SVM as a meta-model to generate output using the base models’ predictions. The proposed model was generalized using the UNBC–McMaster dataset. The findings highlight that the proposed model addressed the existing limitations and produced an exceptional result. The healthcare centers can benefit from the proposed model to identify pain using facial expressions. Internet of Things cameras can be installed to monitor the patients using the proposed model. However, the authors faced a few challenges in the feature extraction process. The lightweight architecture of the ShuffleNet V2 models demanded substantial training for extracting the key features. The fine-tuning is adequate for the CatBoost and XGBoost models to improve the effectiveness of the suggested PIM. Due to their complexity, ensemble models require more substantial computer resources for training and inference. The intricate nature of ensemble models may decrease their interpretability and pose difficulties in deploying them in contexts with limited resources. The effectiveness of ensemble models is dependent upon the heterogeneity and magnitude of the constituent base models. Maximizing the advantages of the proposed model requires ensuring sufficient diversity across base models while avoiding any repetition. The transformer and liquid neural network can be employed to enhance the efficiency of the recommended model.

ACKNOWLEDGMENTS

The authors extend their appreciation to the King Salman Center for Disability Research (funder ID: http://dx.doi.org/10.13039/501100019345) for funding this work through Research Group no. KSRG-2023-349.

REFERENCES


