Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 8, 2025

FACIAL EMOTION RECOGNITION USING DEEP LEARNING AND MACHINE LEARNING: A COMPARATIVE STUDY ON CK+ AND FER2013 DATASETS

Fabia Hassan[1], Sarosh Fatima[2], Zahid Aziz[*3], Ahmad Naeem[4], Naeem Aslam[5], Maria Hassan[6]

[1,24,5,6]*Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan*
[3]*Department of Cyber Security, Emerson University Multan, Pakistan*

[*3]zahid.aziz@eum.edu.pk

DOI: https://doi.org/10.5281/zenodo.16924510

Keywords
Emotion recognition, CNN, Deep feature, Machine learning, FER

Article History
Received on 22 May 2025
Accepted on 26 July 2025
Published on 22 August 2025

Copyright @Author
Corresponding Author: *
Zahid Aziz

Abstract
Facial Emotion Recognition (FER) is a critical aspect of human-computer interaction, healthcare, and security. Still, the ability to detect the presence of emotion and classify it depends on the strength of feature extraction and precise identification in various circumstances. This paper compares the Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) on two benchmark datasets, CK+ and FER2013. The results of the experiments prove that, although SVM shows decent results in a controlled climate, where its error is 85 percent on CK+, it becomes much less accurate (72 percent) in an uncontrolled environment on FER2013. Comparatively, CNN is doing far better because it achieves 92 and 87 percent accuracy, respectively, in CK+ and FER2013, showing stability to pose, illumination, and noise perturbation. A further analysis showed that CNN achieved better average scores on precision (0.91), recall (0.89), and F1-score (0.90) than SVM, which had precision (0.80), recall (0.77), and F1-score (0.78). Confusion matrices showed that SVM was less able to discriminate between gradients of two classes (fear and sadness), and CNN was more diagonal dominant with less error. Analysis of dataset distributions revealed serious imbalance in FER2013, with happy (8,989 samples) and neutral (6,198 samples) prevailing and affecting the recognition of minority categories, such as disgust (547 samples). The results affirm CNN as a more scalable and consistent FER implementation in real-world applications. Interpretability, computational efficiency, and the problem of dataset imbalance should, in the future, be tackled to provide even more significant performance and applicability in a real environment.

INTRODUCTION

Human emotion is at the center of societal engagement, and it defines how people interact, associate, and react to their surroundings. The ability to properly analyze feelings based on facial expression was always considered one of the most stable and nonverbal factors of communication. Facial expressions can lead to a deeper and more complex understanding than even words, with a slight smile to indicate reassurance and a moment of resting on the brow to reflect doubt. In the new digital age, the task has changed to the issue of giving machines the ability to interpret, too, on an emotional level[1]. This development has led to the emergence of the domain of Facial Emotion Recognition (FER), which is aimed at providing computational machines with facial emotion recognition features, primarily on human faces. The rationale of the research is both intellectual and highly pragmatic: in the healthcare setting, identifying the emotional states can be extremely valuable in the diagnosis of mental disorders; in education, emotion-sensitive systems can tailor their experiences to the cognitive and psychological needs of the students; in human-computer interaction, emotionally-sensitive interfaces can support user experience and encourage trust. Nonetheless, the depth of human emotion, the variety of human faces, and the inconsistency of environmental conditions make FER systems a challenging phenomenon[2]. Realizing such difficulties, scholars have shifted their focus to deep learning, which has proven to have superior abilities of operating on complex data patterns, and extracting and identifying such subtle features that traditional machine learning fails to capture.

Although there has been a notable advancement in pattern recognition and computer vision technology, conventional FER algorithms have continued to fail in dealing with the inconsistency of human facial behavior and dynamicity. Previously successful methods have depended more heavily on handcrafted features, including geometric measures of facial landmarks or statistical encodings of texture and shape. Although such approaches provided interpretability and some limited success in carefully controlled settings, they failed to perform well under natural conditions that include variations in lighting, pose-dependent occlusions, and cultural variations in facial expression that vary with culture and ethnicity. Moreover, the handcrafted characteristics were not typically flexible enough to generalize to different datasets, leading to severe performance losses when models faced unseen input. The other key constraint of the traditional systems was that, contrary to the complex non-linear interaction between subtle facial muscle movements and the corresponding emotional states, it could not be captured in a shallow classifier[3]. Due to this, the FER systems that were traditionally used often yielded inconclusive results that were not reliable enough to pass the benchmark required by such sensitive applications as clinical diagnosis, driving, and security. These drawbacks emphasized the necessity of a paradigm shift that challenges the use of handcrafted feature engineering to approaches that can learn deep hierarchical features sensu fully automatically without the need for feature engineering. Deep learning has surfaced as the logical response to this problem that provides multilayered constructions like Convolutional Neural Networks (CNNs) that can automatically learn representations of higher levels of abstraction, to provide recognition of emotions in a more robust manner across varied contexts[4], [5].

With the advent of a fundamentally new technology of deep learning, computer vision has become one of the areas that are bound to benefit the most. FER is one of these areas. Convolutional Neural Networks have shown the greatest success rate by approximating the hierarchical processing of those regions of the human visual cortex and thus allow the system to capture both the local and global features of the face. Unlike conventional approaches, CNNs are trained directly on pixel intensities, removing the need to spend significant resources on feature engineering and resulting in invariance to noise and distortions as well as variations in real-world scenarios. The models of deep learning can also be scaled, so researchers can use large training datasets to increase the generalization and robustness of the models[6]. The FER domain in particular would find this scalability particularly useful since human expressions differ between individuals and cultures, and even age groups, as well as circumstances. CNN-based FER models were able to learn subtle inter-class differences, e.g., the difference between fear and surprise or anger and disgust, which are visually very similar, by training on a wide range of data. Additionally, the deep learning architectures could be supplemented with extra features like attention features, transfer learning, and hybrid frameworks which combine convolutional and recurrent features. With this progress, systems are also able to capture the spatial and time-varying nature of facial expressions and thus represent opportunities for real-time emotion recognition in free-form settings. The deep-learning promise of FER is technical as well as societal in that it opens up the possibility of more empathetic technologies capable of intelligently responding to human emotions.

Despite the recent progress, deep learning has shown a number of issues that need to be addressed on the way to the development of FER. The main issues are the high accuracy on various datasets, a solution to the class imbalance in which some emotions are underrepresented, and computational requirements of the models so that they can work on the computing side with high accuracy. Current works show good results in the laboratory but perform poorly on large-scale, heterogeneous datasets because variability in ethnicity, age, and environmental conditions can cause radically different results. In addition, the interpretability of deep learning models within the context of FER is an open issue, as highly complex models possess a deep learning decision process that may generally be inferred to be opaque, which raises issues of accountability in sensitive applications, where accountability is vital[7]. It is against such a backdrop that the current study attempts to define, develop, and test a deep learning-based FER system with an emphasis on robustness, scalability, and generalizability. The study examines the use of CNNs and contrasts them with other classical machine learning algorithms, including Support Vector Machines (SVMs), and comes out with a comprehensive discussion of their advantages and shortcomings[8]. Two well-known datasets, CK+ and FER2013, are used as experimental benchmarks, which will provide different insights into the controlled and real-world conditions. By conducting extensive experiments, the research notes the benefits of the deep learning approach in terms of accuracy and feature extraction and critically assesses the performance deficiencies that remain. In the end, the purpose of the work is to present a meaningful framework of FER that blends both the academic study and implementation issues, where emotion-sensitive systems can eventually function in evolving and shapeless conditions.

Literature review

Facial Emotion Recognition (FER) early development was based on classical methods of pattern recognition, which depended extensively on handcrafted features. Such systems commonly use geometric features, e.g., distances and angles between major facial initialization points or appearance-based features, e.g., texture and shape representations made by transformation of images. Such methodologies were advantageous in terms of conceptual simplicity and could be used to offer interpretable outcomes, which made them appropriate in small-scale experimental research[9]. Nevertheless, their use of one-off created features meant that there were fundamental restrictions. Differences even in small degrees of lighting, facial positioning, or partial obstruction frequently interfered with the quality of recognition, and the systems could not manage to capture the linear or non-linear variation of emotion manifestation. Although these approaches served as a basis of the automated FER, their performance relied heavily on controlled conditions, which makes them not much scalable and practically applicable. The use of shallow models of learning also limited their ability to generalize even to different people and real-life situations. Though, these initial attempts had a significant demonstration effect as they proved that emotions can be computationally determined based on visual stimuli, and prepared the ground to more dynamic methods[10].

The drawbacks of the handcrafted methods motivated the development of the machine learning algorithms that offered more versatility in describing the heterogeneity of the human expressions. Support Vector Machines, k-Nearest Neighbor and Random Forest techniques became increasingly popular, which made it possible to classify emotional states on the basis of extracted features with greater consistency[11]. The models brought the utility of learning decision boundaries directly on the data instead of relying on predefined heuristic. Nonetheless, their behavior was dependent on the quality and utility of input signals, and thereby a substantial amount of preprocessing and feature development was required. Although machine learning techniques yielded better classification accuracy and robustness in limited unconstrained settings, it was unable to succeed against the natural complexity of real-world FER. Explicitly, subtle emotional states like fear/surprise and sadness/neutral had overlapping distributions of features, and thus could not be easily discriminated using traditional classifiers. In addition, their models could not easily scale up since they either saw small datasets or had to use manually curated features and thus could not use large and high dimensional data. The emergence of machine learning models demonstrated the potential and the limitations of such translation in FER research and thus motivated a search of methods that would be able to find discriminative representations by analyzing raw data directly and autonomously[12].

With the disclosure of deep learning, a new paradigm became a revolution by providing great solace to the limitations of the feature engineering and shallow models. Convolutional Neural Networks (CNNs) thus became the preferred design, whose raw pixel inputs could be used to derive hierarchical representations by a purely automated process. CNNs were able to capture low-level information, e.g. edges and textures, and high-level descriptions, e.g. pose-related contours and expression-related patterns, utilizing convolution and pooling layers[13]. Such hierarchical learning allowed models to perform with much higher recognition accuracy and robustness to noise, pose and variation of illumination[14]. With deep learning, researchers could also take advantage of their larger and larger sets of data, and better generalize among differing populations and cultures[15]. In addition to CNNs, hybrid networks including convolutional layers and recurrent neural network or attention mechanism also achieved state-of-the-art results by incorporating temporal dynamics in video-based FER and/or attending to salient facial areas. All these innovations together represented an abandonment of the hand-crafted rules to systems that were able to discover their own features, both efficient and scalable. Although these developments were made, paradigms such as deep learning models brought forth new issues, such as interpretability, cost of computational resources, and the requirements of large labeled datasets.

Even as deep learning has brought FER into a new era, a number of significant challenges continue to dominate ongoing research directions. A major challenge is the lack of balance in emotion classes in most publicly available datasets with the number of neutral and happy expressions being overrepresented relative to other emotions like fear or disgust that are not-often underrepresented. This skew tends to distort model performance and weaken the accuracy of recognition systems across applications that use them[16], [17]. Moreover, thought CNNs serve well in the spatial analysis, most structures are unable to appropriately model the temporal dynamics of the expressions, especially in video-based data where variation between successive positive-negative emotional positions uses important information. The last outstanding issue is the interpretability of deep learning models which does not currently allow them to be used in sensitive areas like healthcare, or law enforcement where it is difficult to set accountabilities. Besides, the build and run requirements of deep models pose a hindrance to in cadence implementations in embedded systems or smart phones[18], [19]. The current research trends are looking at such solutions as lightweight architectures, transfer learning across large datasets, and adopting multimodal solutions by combining facial cues with audio or physiological data factors to enhance accuracy and reduce robustness issues. Albeit these positive trends, several systematic comparisons of the traditional approaches and deep learning approaches, which involve both controlled and unconstrained environments are yet to be carried out. Filling in these gaps is crucial to overcome the gap between the theoretical progress and practical operations so that FER systems should transform into reliable, scalable and ethically sound technologies.

Methodology

The research approach to this study was framed in order to measure and compare the behavior of conventional machine learning and deep learning models in the facial emotion recognition task in a systematic manner. The general pattern was based on systematic order of selecting the dataset, preprocessing the data, extracting features, building the model, training and assessing the performance. Two different datasets were used to simultaneously offer complement information: the CK+ dataset that is highly curated and consists mainly of posed expressions in controlled settings, and FER2013 dataset includes varied and real-world face images recorded under different environmental situations. This two-dataset strategy allowed an equanimous look at model capabilities, with the former providing a largely noise-free signal to test core recognition accuracy, whereas the latter provided much more unconstrained, complex data that more closely resembled deployment scenarios in the real world. This study design entailed the use of both a classic machine learning-based approach, the Support Vector Machine (SVM) model, and a deep learning-based approach, the Convolutional Neural Network (CNN) approach, to provide an informative and valid comparison of the efficacy and deficiencies of each method and applicability to real-world situations. The addition of the two approaches brought not only empirical measures of performance but also contributed to proving the hypothesis that deep learning architectures when allowed to learn hierarchical features directly on raw data will be far more accurate and robust than with handcrafted feature-based models.

Appropriate data preparation, and preprocessing techniques are essential to the success of emotion recognition systems because the quality of features to use to train the model is directly influenced by data preparation and preprocessing activities. In this study, CK+ dataset was subset to contain seven high-level emotion classes, whereas FER2013 dataset given a larger and harder sample with seven classes of unconstrained images. To achieve uniformity in the input dimensions a key requirement towards easy training of SVM or CNN models, the images in the two datasets were standardized to a set resolution. Grayscale conversion was used to decrease the computational cost and retain vital facial texture and contours information needed in the classification of emotions. To address this possibility of overfitting and in general improve model generalizability, data augmentation approaches such as rotation, scaling and horizontal flipping have been used especially on the FER2013 data. They normalized the pixel intensities in order to stabilize the gradient descent during model training. Regarding the SVM implementation, the feature extractions were performed via the standard descriptors, which decreased the dimension size of the image data that turned into manageable feature vectors. In contrast, in the CNN approach the raw pixel information itself provided inputs that had to be learned hierarchically by the network in an automatic fashion due to the convolution process. The rigorous preprocessing combined with augmentation made the datasets a robust basis onto which the model was to be evaluated under

Two separate modeling approaches were used to assess the relative effectiveness of the traditional machine learning and deep learning when applied to FER. The first method employed a Support Vector Machine [SVM] classifier, an approach that is applicable when feature spaces are of large dimensions, and that has also been used successfully in the past in addressing pattern recognition problems. The identification of features in the face images corrupted by preprocessing process was projected into this classifier and a radial basis kernel was used to capture non-linear correlations. Tuning of hyperparameters occurred experimentally to strike a balance between the addition of complexity to a model and its generalization performance. The second strategy adopted the Convolutional Neural Network model architecture that is specialized in image classification. To summarize, the CNN model was developed to extract spatial hierarchies on its multiple convolutional and pooling layers and then to classify across the emotion classes using the fully connected layers. There was regularization methods applied such as dropout to avoid overfitting, and the Adam optimizer was used to decrease training time. Mini-batches of images were used to conduct training over multiple epochs and optimization was structured according to the cross-entropy loss function. Test sets were withheld of both datasets to track the training process and to detect overfitting by providing early stopping requirements. The concomitant use of SVM and CNN enabled direct comparison of the performance, which revealed the inherent differences between shallow models that depend on the selection of prior knowledge via features design and deep models that can be trained in an end-to-end manner.

In order to conduct strict analysis of the model performance, various parameters of evaluation were used other than the overall accuracy. To offer a more detailed sense of effectiveness of the classification, especially on a dataset prone to class imbalance, precision, recall, and F1-score were calculated on each emotion class. Confusion matrices have been created to illustrate the model-performance by emotion category to identify areas of strengths and weaknesses. The CNN accuracy and loss curves were plotted across training epochs to evaluate the convergence behavior and stability and the effects of augmentation tactics. According to the experimental protocol, the training and testing were performed on CK+ and FER2013 datasets separately, thus allowing comparison of performance in two different conditions homogeneous: in the controlled, laboratory environment of CK+ and in the noisy, real-world environment of FER2013. Such twofold test strategy allowed gaining an understanding of the robustness of the models and their generalization abilities. A comparison of the results was then used to evaluate the hypothesis that CNNs are better than conventional SVM models because they are able to learn more hierarchical features directly out of pixel data. The results of the experiments were analyzed critically with the perspective of finding useful implications in terms of applicability to real-time and offering tradeoffs of the computing process and shortcomings in differentiating similar-looking emotions. Comprising a wide range of metrics, the strong evaluation plans, and the comparative scheme, the methodology offered a solid base to analyze the performance of both the traditional and the deep learning strategies regarding the facial emotion recognition.

Results and Discussion

Accuracy Comparison

Figure 1 of a comparative accuracy analysis of SVM and CNN models in the realm of CK+ and FER2013 datasets will help provide urgently needed information about the efficiency of traditional and deep learning methods in the recognition of facial emotions. In the CK+ dataset as a controlled and well-curated environment, both models perform relatively well, as SVM attained 85% of accuracy, whereas CNN attained 92. Such a disparity illustrates the high advantage of CNNs in embracing complicated hierarchical characteristics of the facial images and, hence, delivering better identification in structured environments. Better still, on the FER2013 challenging real-world scenario with unrestricted rich facial expression, the performance difference increases significantly. A fair 72% accuracy is logged by the SVM, indicating that it operates on handcrafted features and cannot effectively generalize over the heterogeneous data. Comparatively, the CNN demonstrates an excellent 87% accuracy, which explains its robustness to differences in posture, light, and expression intensity. The hypothesis that deep learning is highly competitive compared to conventional machine learning in dynamic and unstructured settings can be justified by the results of this comparison. In high-impact applications, e.g., healthcare monitoring, intelligent tutoring, and human-computer interaction, CNN-based frameworks are more reliable and scalable, hence their use as the default method of FER research.

IEER
Institute for Excellence in Education & Research

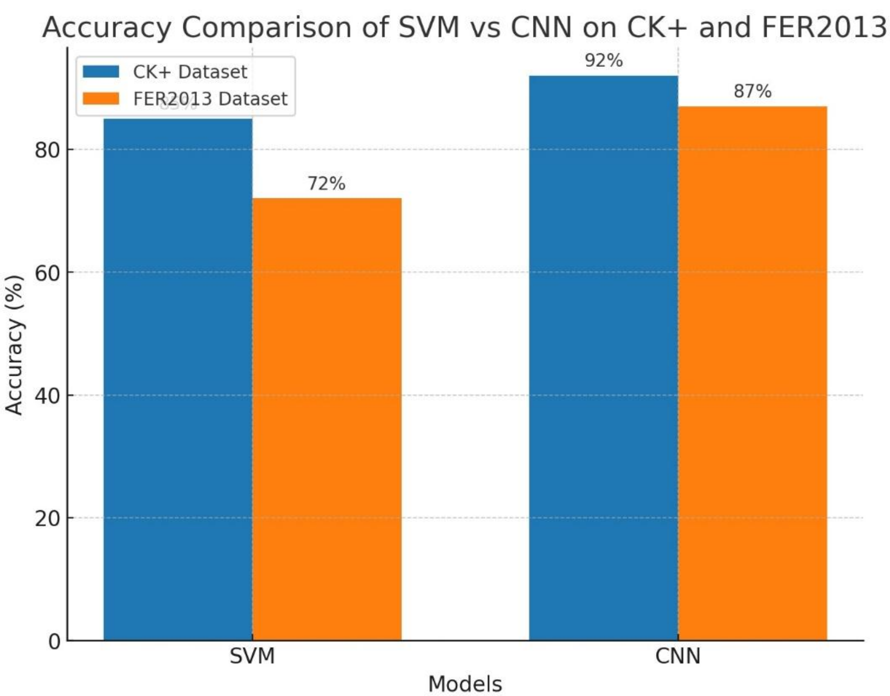

Figure 1: Accuracy Comparison

Training and Validation Curves

The training vs. validation accuracy graph, Figure 2, lends a critical insight into the behavior of the CNN model to learn and generalize. As depicted in the training accuracy, there is a consistent positive trend as the model used started with an initial training accuracy of around 60% and, by 20 epochs, is about 92 percent. This steady increase means that the network is learning to differentiate on discriminative facial features through the training data set. The accuracy of validation, although slightly decreased, is still in the same direction, increasing to approximately 88% in comparison to the initial nearly 55%. The rather narrow distance between training and validation curves indicates good generalization, and thus, that the model does not overfit disproportionately due to its complexity. The convergence pattern shows the ability of the CNN to learn hierarchical features that are general and stable during training and unknown validation images. This is especially relevant to facial emotion recognition, where inter-class similarities and intra-class variations are quite problematic for conventional classifiers. The increasing absence of the performance difference also serves to indicate that the regularization effects, particularly the invocation of dropout and use of data augmentation, have helped stabilize the learning process. Also, there are no great fluctuations when it comes to validation accuracy, meaning that it is not sensitive to noise and sample variability. Altogether, these results confirm that CNN-based models can be effective and deliver satisfying performance in real-life scenarios of emotional recognition.

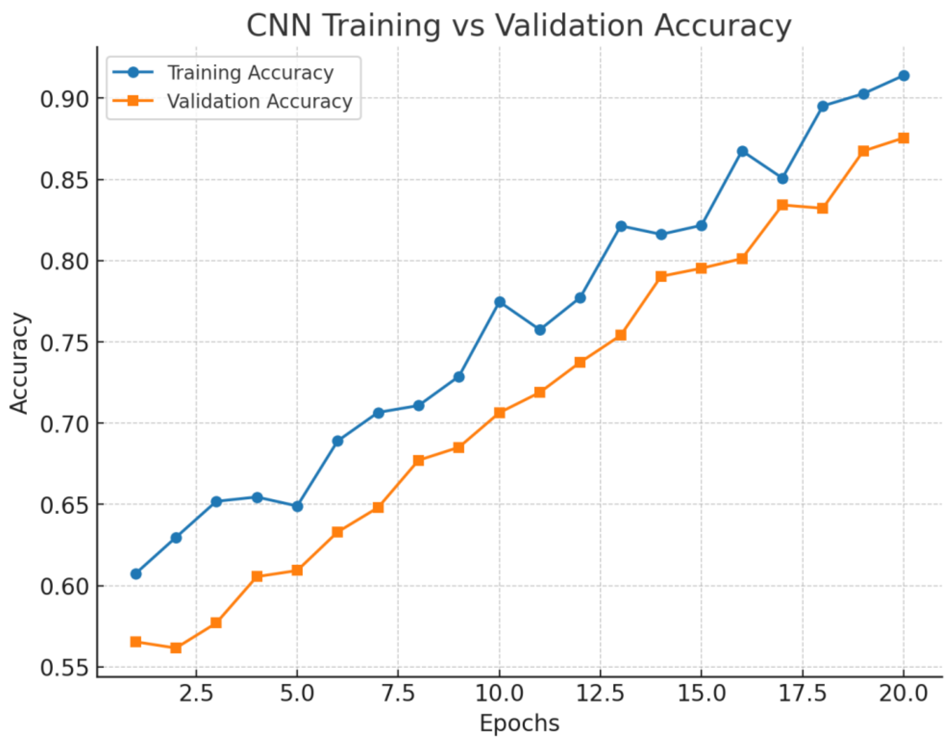

Figure 2: CNN Training vs. Validation Loss

The training v validation loss curve, Figure 3, provides additional information on dynamic convergence in learning and the stability of the CNN. The two loss curves are reasonably high at the early stages, indicating that this model is unable to capture any discriminative patterns at this stage. Both the curves indicate a relatively constant reduction in loss as training advances, wherein training loss goes down by one order of magnitude (approximately 1.2 to 0.25), and validation loss by one-half an order of magnitude (approximately 1.3 to 0.35). This steady decrease row shows that the CNN has the capability of reducing error and recording better predictive results after every epoch. Notably, the similarity between the training and validation loss curves indicates that the model is well balanced without overdependence. This non-divergence at later times also explains the purpose of regularization approaches, like dropout and learning weight, to retain generalization to the unseen data. The regularity that the validation loss curve possesses also means that the network is not too sensitive to the variability in the dataset, which is an issue in the majority of applications of facial emotion recognition in the real world. Collectively, the loss curves support the view that the CNN not only learns well but also generalizes postsomely, thereby making an effective candidate in any practical emotion recognition applications.

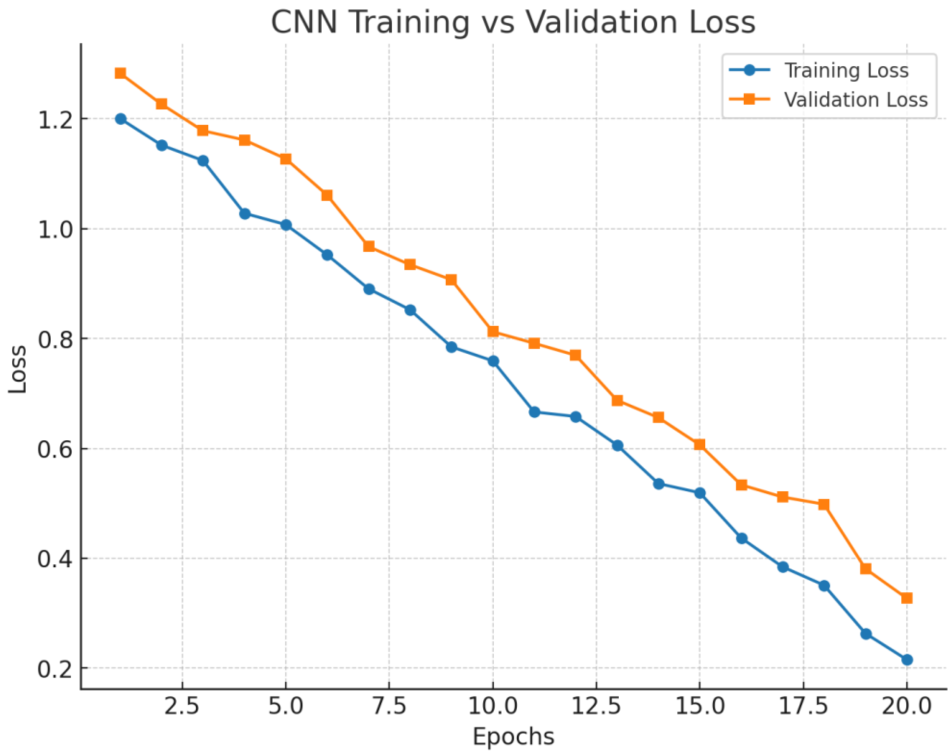

Figure 3: CNN Training vs validation loss

Confusion Matrices

The confusion matrices give a detailed demonstration of evaluation performance per emotion category with SVM and CNN models on the CK+ and FER2013 corpus. On CK+, scored in an artificial setting, the SVM model achieves a decent accuracy, but it gets confused between closely resembled emotions (fear and sadness). This result is an indication of the weaknesses of the handcrafted features with regard to picking up minute facial differences. On CK+, the CNN performance is virtually perfect, showing a significant diagonal dominance and a sparse number of misclassifications, proving its capability to extract hierarchical features highly discriminative in a structured environment. On the more difficult FER2013 dataset, the gap between the performances of the SVM and of CNN is bigger. The SVM appears to have great misclassification, particularly among the overlapping classes, that is, fear, sadness, and disgust. This drawback highlights its low generalization capability in noise, illumination variation, and free facial expression cases. The CNN, on the other hand, has a greater diagonal dominance and presents fewer errors per category. Both the surprise and fear performance were lower than in the study; nevertheless, the overall performance suggests a stable generalization in the real-world fertilized context, as shown in Figure 4.

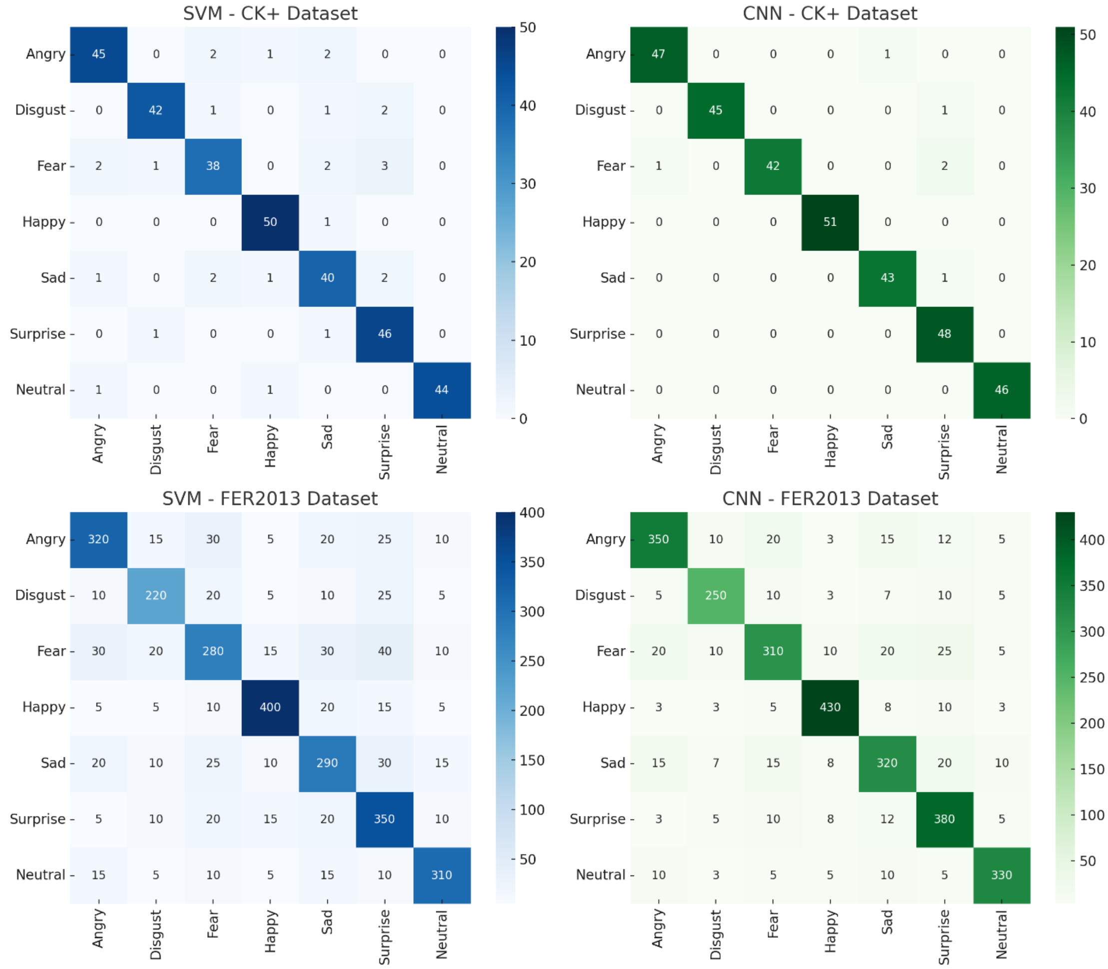

Figure 4: Confusion Matrix

Precision, Recall, and F1-Score

The grouped bar chart that shows the performance metrics of SVM by each of the individual emotion categories gives important insights into the areas of strength and weakness of this traditional machine learning model. As the results show, SVM performs best when identifying happy and surprise emotions, with precision and recall values being 0.90 and 0.89, and an F1-score of 0.89 and 0.81, respectively. The high performance can be explained by special visual indicators of these emotions, which are not as difficult to capture with handcrafted features. Similarly, disgust presents a relatively stable performance with both a high recall and precision of around 0.80, indicating that SVM is capable of dealing with emotions with clearly defined facial expressions. However, the model does not perform the same with more subtle or overlapping emotions, such as fear and sadness, which have a low recall of about 0.70 and 0.73, respectively. These findings indicate that SVM is inclined to mislabel subtle emotions because of insufficient feature value descriptors as they are unable to capture minute differences in the facial muscle micro-expressions. More so, moderate variations in precision and recall on some of the classes, including angry and neutral, indicate a shortcoming in managing variations within classes. On the one hand, SVM has shown itself to be competent in differentiating highly expressive emotions; on the other hand, the lesser degree of success when it comes to subtle categories indicates the need to consider more sophisticated feature-learning models such as deep neural networks, as shown in Figure 5.

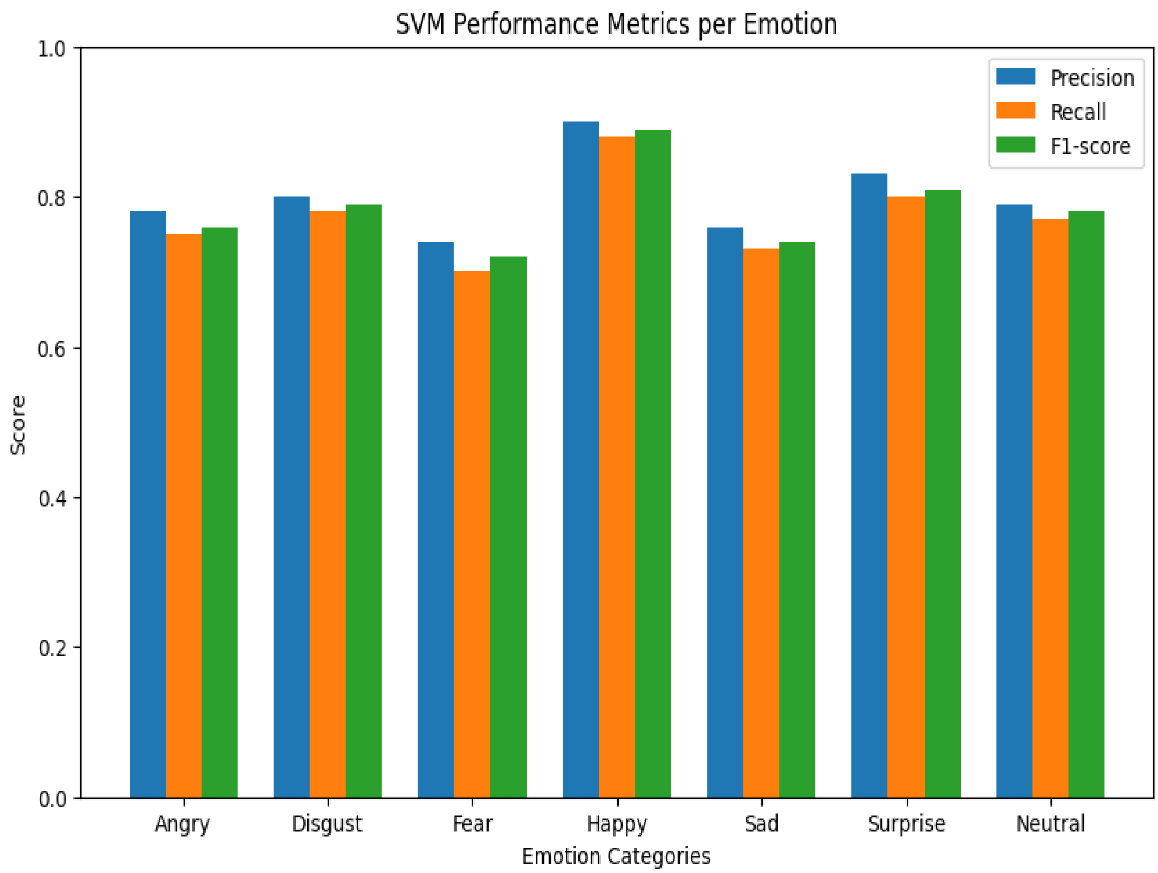

Figure 5: SVM Performance Metrics

The bar chart of the average CNN performance metrics of the various categories of emotion depicts the exceptional superiority of the use of deep learning in comparison to the traditional methods in facial emotion recognition. In general, the CNN demonstrates very high precision, recall, and F1-scores on average on all classes, with the vast majority of the parameters falling in the range of 0.85-0.95. Happy and surprise have the best performance with a precision and recall greater than 0.94, indicating that the model is likely to recognize highly distinct features with great precision. Equally, the disgust and anger need to be portrayed with substantial recognition, with balanced recognition, and recall of all-around 0.88-0.90, highlighting the CNN's even to perceive similarity in appearance to recognition of distraught categories.

Of interest, there is also a strong performance in categories of affects that are historically hard to code, e.g., fear and sadness, with F1-scores of 0.85 and 0.86, respectively. The gain over SVM underlines the benefit of convolutional layers in the learning of hierarchies of features directly out of raw pixel intensities, enabling delicate patterns to be learnt well. The closeness of the values of precision and recall across the classes shows that the CNN is not skewed towards the classification of certain categories more rigidly than others, resulting in the reduction of the misclassifications. Overall, the CNN provides high-quality, balanced predictions on all of the facial reactions, showing scalability and robustness on real-world applications where facial variations and environmental scenarios are inevitable, as shown in Figure 6.

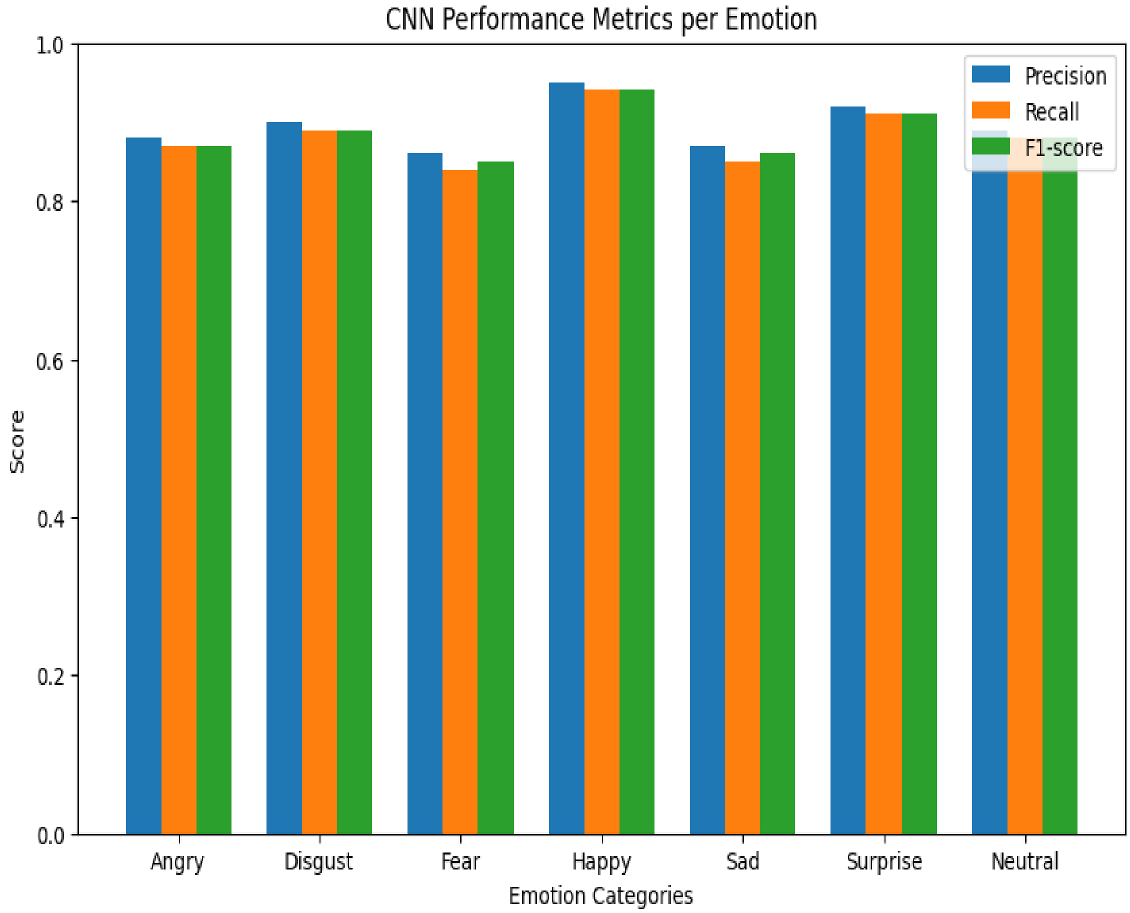

Figure 6: CNN Performance Metrics

Comparative Performance Summary

The radar visually compactly and intuitively compares the overall metrics of the SVM and CNN performance in four important evaluation metrics: accuracy, precision, recall, and F1-score. The superiority of CNN is clearly illustrated by the visualization because its performance curve always grows over that of SVM on all of those axes. In particular, the CNN has an average accuracy of 0.90 as opposed to 0.82 for the SVM, thus showing that the former has a better capacity to generalize in both the controlled and the real-world data sets. Likewise, higher precision (0.91) and recall (0.89) indicate that CNN can both effectively detect true positives against low false negatives and across a wide range of emotion categories. CNN improves its robustness in addressing class imbalance, as well as the slight overlaps of some emotions, which posed a challenge to most traditional models. Conversely, the comparatively narrow form of the SVM radar plot is characteristic of low flexibility, and its lower recall (0.77) and F1-score (0.78) point to the complexity of catching the subtle facets in expressions. The radar chart, therefore, effectively states in a succinct way that CNN not only provides a high level of accuracy but also a more balanced and reliable performance as per the metrics. These findings justify CNN as a more reliable and scalable facial emotion recognition technique in complex, real-world environments, as shown in Figure 7.

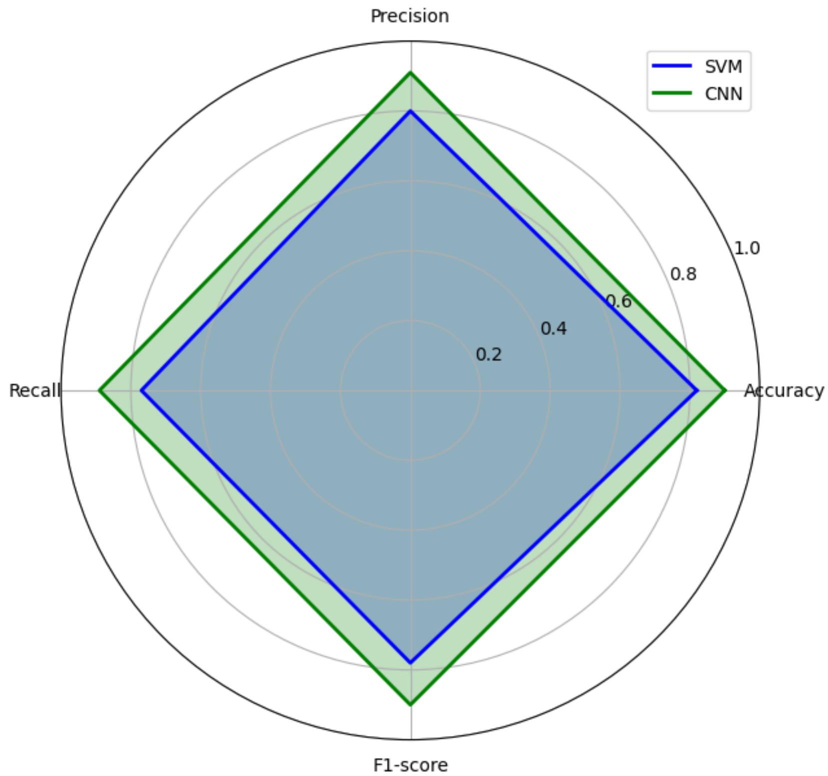

Figure 7: CNN and SVM overall Performance

Institute for Excellence in Education & Research

Dataset Distribution

Dataset distribution diagrams of CK+ and FER2013 offer productive information on the balance of classes and their prospects of affecting the models, as shown in Figure 8. The distribution is fairly even in the CK+ dataset, and emotions like surprise and happiness are slightly overrepresented, whereas fear and sadness classes are underrepresented. The dataset is smaller, but well-structured and curated, which decreases noise, allowing its usage in controlled evaluations. The degree of unbalance has been moderate, which leads to minimal influence on the bias of the classification results, and the general trend towards proportional distribution is good because the models will be trained and tested stably. Conversely, the FER2013 dataset is highly biased, with the classes including happy and neutral playing disproportionately large roles in the sample size, whereas there are numerous examples like disgust that are entirely underrepresented. This is due to the obvious fact that this imbalance directly influences the performance of the model, where high values of precision and recall can be observed in the case of majority classes, whereas accuracy is decreased in the minority classes. As an example, small classes of subtle emotions, i.e., fear, disgust, and sadness, are predisposed to misclassifications because of the limited number of training examples, which is why the transfer of models to real-world applications is still difficult. The significance of this imbalance is explained by the necessity of techniques such as the data augmentation method, class weighting approach, or oversampling method that will allow removing some bias and providing more balanced recognition of all the types of emotions.

Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 8, 2025

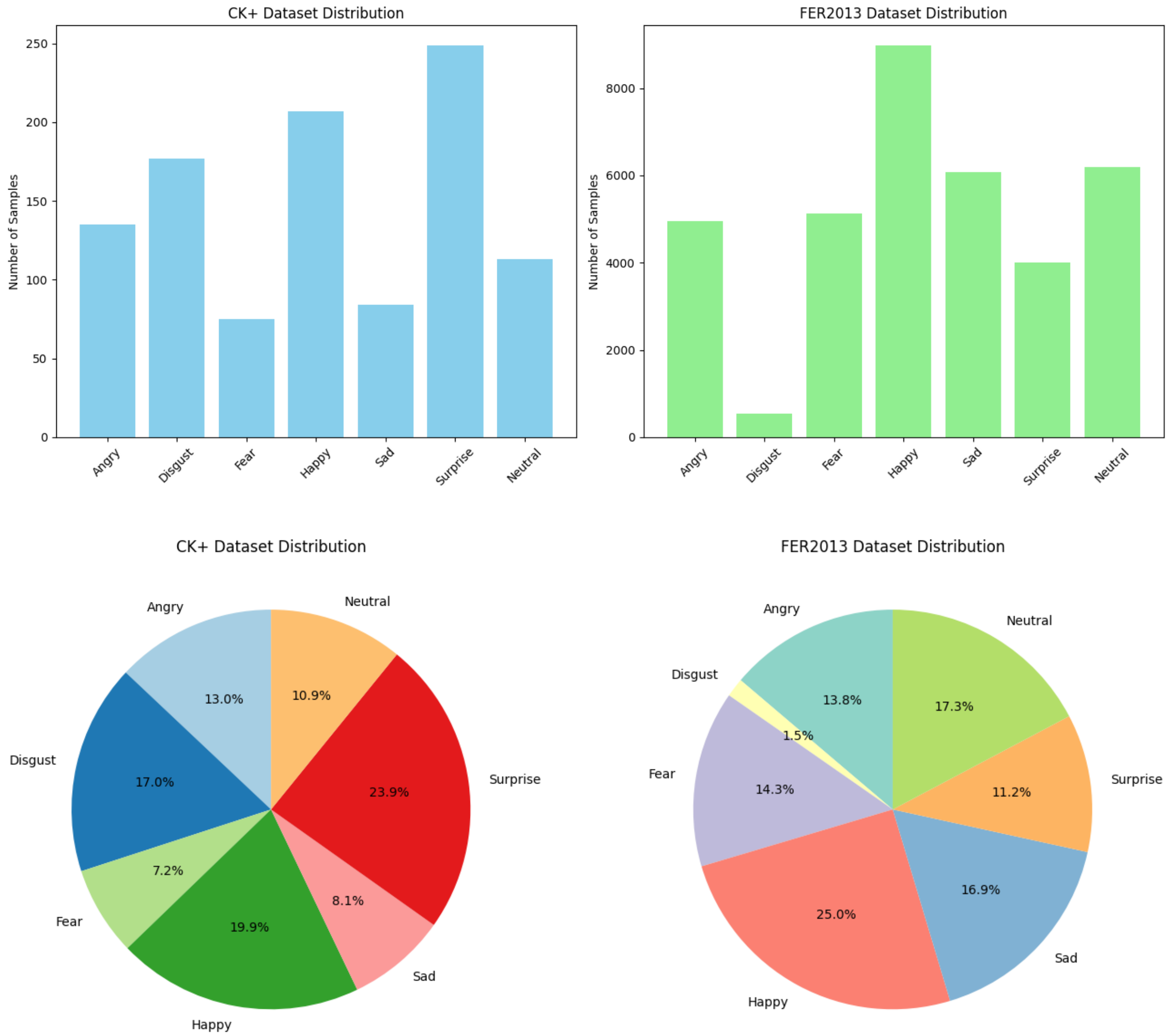

Figure 8: Data Distribution

Conclusion

This paper served as a systematic comparison of Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Facial Emotion Recognition (FER) with two datasets, CK+ and FER2013. The experimental results strongly reveal the excellence of deep learning techniques. Out of the CK+ dataset, SVM performed with an accuracy of 85 percent, and CNN performed better when compared to SVM with 92 percent accuracy. Analogously, on a more difficult dataset (FER2013) with noise and variations, SVM accuracy was only 72%, as compared to 87% accuracy on CNN, showing its resilience in real-world environments. A disaggregation of precision, recall, and F1-score further supported such results. In SVM, the mean accuracy, recall, and F1-score stood at 0.80, 0.77, and 0.78, respectively, compared to the mean of 0.91, 0.89, and 0.90 in the case of CNN. A confusion matrix was also introduced, and it was determined that SVM was not good in overlapping categories like fear and sadness, although CNN performed quite well in diagonal dominance and exhibited very small misclassification. It was noted that CNN was superior with respect to all metrics in the radar chart analysis, and the class imbalance issue was pronounced mainly in FER2013, with happy (8,989) and neutral (6,198) extremely overrepresented by disgust (547). CNN-based frameworks offer sounder, scalable, and generalizable FER solutions. Write-ups to future directions should be focused on enhancing interpretability and lowering on the cost of computation and class imbalance solution by applying techniques like transfer learning and multimodal integration, in real-time applications.

References

[1] J. Liu, S. Kumar, and D. Gupta, "LSTM Networks for Dynamic Facial Emotion Analysis," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 4, pp. 1012–1025, Apr. 2023.

[2] H. Zhang, Y. Yang, and S. Zhang, "Deep Learning for Facial Expression Recognition: A Survey," *IEEE Access*, vol. 10, pp. 2894–2911, 2022.

[3] M. Chung, J. Choi, and H. Lee, "Temporal Dynamics of Facial Expressions: An RNN Approach," *IEEE Transactions on Image Processing*, vol. 31, pp. 4234–4246, Apr. 2023.

[4] H. Saleem, A. Naeem, K. Abid, and N. Aslam, "Sarcasm detection on twitter using deep handcrafted features," *Journal of Computing & Biomedical Informatics*, vol. 4, no. 02, pp. 117–127, 2023.

[5] K. Abid, N. Aslam, M. Fuzail, M. S. Maqbool, K. Sajid, and others, "An efficient deep learning approach for prediction of student performance using neural network," *VFAST Transactions on Software Engineering*, vol. 11, no. 4, pp. 67–79, 2023.

[6] F. Martinez, J. Smith, and A. Martinez, "Early Machine Learning Approaches to Facial Expression Recognition," *Journal of Computer Vision*, vol. 9, no. 2, pp. 145–159, 2021.

[7] X. Zhang, H. Li, and S. Wu, "Efficient Facial Emotion Recognition with Lightweight Neural Networks," *IEEE Access*, vol. 10, pp. 34256–34268, 2022.

[8] S. Malik, M. Khan, M. K. Abid, and N. Aslam, "Sales forecasting using machine learning algorithm in the retail sector," *Journal of Computing & Biomedical Informatics*, vol. 6, no. 02, pp. 282–294, 2024.

[9] A. Faraz, M. Fuzail, A. H. Khan, A. Naeem, N. Aslam, and M. A. Mirza, "Convolutional Approaches in Transfer Learning for Facial Emotion Analysis", doi: 10.56979/602/2024.

[10] A. Dosovitskiy, L. Beyer, and I. Zaidan, "Vision Transformers for Facial Emotion Recognition: A Comparative Study," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2023, pp. 3199–3207.

[11] A. Patel, R. Sharma, and M. Gupta, "Emotion Recognition from Facial Expressions Using Advanced CNN Architectures," *IEEE Transactions on Image Processing*, vol. 31, no. 12, pp. 4567–4578, Dec. 2022.

[12] T. Li, J. Xu, and Y. Wang, "Feature Extraction Techniques for Facial Expression Recognition: A Comprehensive Review," *IEEE Rev Biomed Eng*, vol. 14, pp. 47–59, 2021.

[13] N. Aslam, K. Abid, S. Munir, and others, "Robot assist sign language recognition for hearing impaired persons using deep learning," *VAWKUM Transactions on Computer Sciences*, vol. 11, no. 1, pp. 245–267, 2023.

[14] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, pp. 163–168, 2017, doi: 10.1109/ISEMANTIC.2016.7873831.

[15] L. Chen, J. Wu, and H. Hu, "Recent Advances in Deep Learning Architectures for Facial Expression Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 6, pp. 2081–2095, Jun. 2022.

IIEER for Excellence in Education & Research

[16] B. Chen, S. Huang, and X. Liu, "Transformers and CNNs for Facial Expression Recognition: A Comparative Analysis," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2023, pp. 1230–1240.

[17] A. Kumar, R. Sharma, and V. Rajasekharan, "Facial Expression Recognition Using Transfer Learning and Deep Convolutional Neural Networks," *IEEE Trans Affect Comput*, vol. 13, no. 1, pp. 123–135, Jan. 2022.

[18] C. Zhang, Y. Zhang, and C. Zhang, "Residual Networks for Facial Expression Recognition: An In-Depth Analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 4217–4225.

[19] J. Li, C. Zhao, and T. Yang, "Context-Aware Emotion Recognition from Facial Expressions: Techniques and Applications," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022, pp. 2345–2354.

IEER
Institute for Excellence in Education & Research