Rule-based Emotion Recognition

Shamanthak Hegde, Sriranjan Srikanth, Sujith Tellakulla, Suprad Parashar, Venkat Nikhil
Arizona State University
Tempe, AZ, USA, 85281
{shegde23, ssrika21, stellak1, sparash8, vmangip1}@asu.edu

***Abstract*—This paper introduces an innovative approach to emotion recognition through a rule-based classification paradigm grounded in abductive reasoning. Our method integrates the facial feature extraction capabilities of the Mediapipe library with the rule-learning prowess of the RIPPER algorithm, aiming to efficiently and transparently classify emotions within facial images. Leveraging the spatial landmarks and temporal sequences provided by Mediapipe, our methodology establishes a comprehensive foundation for subsequent rule generation. The RIPPER algorithm, applied as the core of our approach, distills concise and interpretable rules from the intricate relationship between facial features and emotion labels. Through extensive experimentation on a diverse dataset, we showcase the system's competitive accuracy, emphasizing the interpretability and efficiency of our rule-based model. The results underscore the potential of our method in real-world applications such as human-computer interaction, mental health monitoring, and affective computing. This work not only advances the field of emotion recognition but also lays the groundwork for interpretable machine learning systems in emotionally sensitive domains.**

Fig. 1. Examples of Image with 'Happy' and 'Fear' Emotion, with it's annotation of facial features

I. Introduction

The demand for emotionally intelligent computing systems has spurred advancements in emotion recognition, a critical component in applications such as virtual reality, human-robot interaction, and healthcare. Despite the success of deep learning in capturing intricate patterns, its inherent lack of interpretability often hinders its application, particularly in contexts where transparent decision-making is essential.

This paper is motivated by the dual objectives of improving precision and interpretability in emotion recognition systems. In addition to leveraging the robustness of Google's MediaPipe library for facial feature extraction, we propose a fusion with the rules generated by the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm. This hybrid approach aims to capitalize on the discriminative power of facial features while providing transparent and human-readable rules for emotion classification.

We extend beyond traditional convolutional neural networks (CNN) by incorporating neurosymbolic learning principles. Rule-based classification, a hallmark of our approach, aligns with the interpretability needs of real-world applications. Our contribution lies in the seamless integration of two key components: the accurate extraction of facial features using Google's MediaPipe and the dynamic rule generation facilitated by the RIPPER algorithm. The latter, known for its ability to produce concise and accurate rule sets through iterative refinement, forms the backbone of our rule-based emotion recognition system. To achieve this, we initially employ the MediaPipe library, renowned for its proficiency in effectively identifying key facial landmarks such as face contours, eye shapes, eyebrow dynamics, and mouth configurations [2]. This library is instrumental in producing a wealth of facial attributes, each associated with corresponding numerical values, thereby capturing the essential intricacies of facial features. RIPPER, a powerful data mining and machine learning algorithm, enhances the existing methodology by extracting valuable insights and patterns from our facial feature dataset. It brings an added layer of sophistication to our research, allowing for a more nuanced and comprehensive understanding of the emotional expressions depicted in facial features.

Figures 1(a) and 1(b) depict a woman expressing happiness and fear. When these images are processed through Google's MediaPipe library, the outputs are annotated images that delineate the various facial features associated with the inputs. These annotations are visually represented in Figures 1(c) and 1(d). Fig. 2 depicts the prominence values for each facial feature in an image where the emotion 'Happy' is expressed.

Through this research, we aim to contribute to the evolving field of emotion recognition, offering a transparent and efficient framework for classifying emotions based on facial features, while simultaneously incorporating the interpretability and adaptability advantages inherent in rule-based classification and neurosymbolic learning.

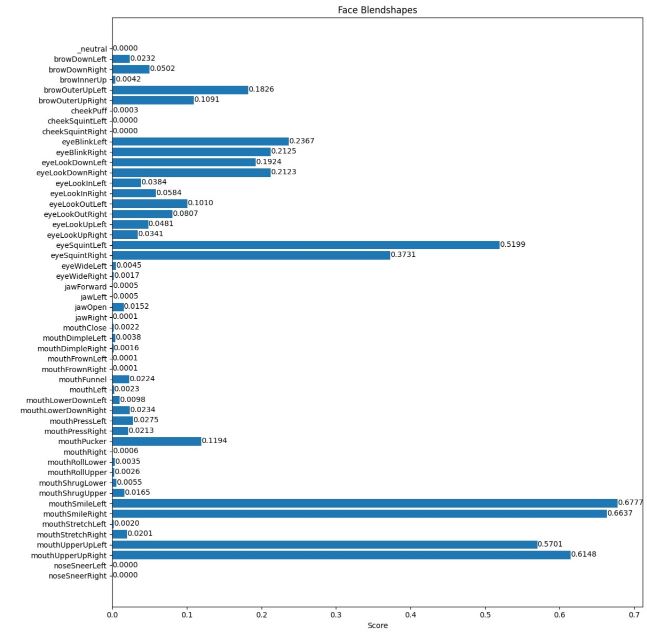

Fig. 2. Prominence Values for each feature extracted by MediaPipe for a 'Happy' image

II. Related Works

A. Emotion Detection and Characterization using Facial Features

This work [4] employed a multi-pronged approach, wherein the training data underwent numerous filtering processes before being characterized using a Support Vector Machine (SVM) optimized by Grid Search. Moreover, the study explored various enhancement techniques for better emotion classification, including the utilization of the Gabor filter, Histogram of Oriented Gradients (HOG), and Discrete Wavelet Transform (DWT). Their findings highlighted that leveraging the HOG, followed by SVM characterization, yielded the highest precision in emotion classification, with an impressive average precision rate of 85%. This work underscores the potential of sophisticated image processing techniques combined with machine learning models in the domain of emotion detection and characterization.

B. Emotion recognition through facial expression analysis based on a neurofuzzy network

In this work [3], the authors study delving into emotion extraction from facial expressions utilized Facial Animation Parameters (FAPs) as per the ISO MPEG-4 standard. This research, through a robust facial analysis system, determines FAPs with corresponding confidence measures. A distinguishing feature is the introduction of a neurofuzzy system that adapts to individual user facial expression nuances. By leveraging clustering analyses of FAP values, the system enables continual learning in real-world scenarios. Preliminary trials with the EC IST ERMIS project datasets validate the efficacy and potential of these technologies.

C. Emotion recognition from text using semantic labels and separable mixture models

This work [7] has unveiled an innovative methodology for emotion recognition from text. The process commences with the formulation of emotion generation rules (EGRs) sourced from psychological principles, providing a blueprint for emotion origination. Textual emotion is represented using a dual system of semantic labels (SLs) and attributes (ATTs), with SLs standing as domain-agnostic elements and ATTs being domain-specific. Employing the Apriori algorithm, the emotion association rules (EARs) for each emotional state are deduced from an emotional text corpus. A pivotal component of this approach is the utilization of a separable mixture model (SMM) to ascertain the emotional resonance of an input text with preset EARs. Notably, the study concentrated on a dialogue system tailored for student expressions, limiting its emotion palette to happy, unhappy, and neutral. Preliminary evaluations showcased the model's potential and its adaptability across diverse domains.

III. Methodology

This work proposes an amalgamation of feature extraction using the MediaPipe library and classification using the RIPPER algorithm as shown in Fig. 3 and Fig. 4. The research framework comprises two primary phases: Training and Testing, each involving several nuanced steps to ensure a robust and comprehensive approach.

The integration of facial feature extraction, rule-based classification, and coverage analysis contributes to a nuanced understanding of emotional expressions in images. However, the experimental nature of certain components underscores the dynamic and iterative nature of research. Future work should focus on refining and expanding these methodologies to enhance the accuracy and generalizability of the emotion classification system.

A. Preliminaries

1) MediaPipe Library: MediaPipe is an open-source framework developed by Google, offering a versatile toolkit for building perceptual computing applications. [6] These applications often involve tasks such as computer vision and machine learning. The framework is designed to streamline the development of real-time applications that process and understand multimedia data, such as images and videos.

The core concept of MediaPipe revolves around constructing a pipeline. This pipeline consists of modular components, each dedicated to a specific task within the application. Examples of tasks include face detection, hand tracking, and pose estimation. These components work together cohesively, allowing developers to build complex applications by selecting and configuring the components that best suit their needs.

MediaPipe provides a set of pre-built components optimized for various tasks, encompassing image processing modules, feature extraction tools, and machine learning models. Developers can customize the pipeline by integrating these components or creating their own, fostering flexibility in application development.

In terms of data flow, the pipeline processes input data, such as video frames or image sequences. This data is sequentially passed through the different components, where each component takes the output of the previous one as input, processes it, and produces a new output. This modular approach facilitates efficient and real-time processing, with optimizations for different hardware, including GPUs.

MediaPipe seamlessly integrates with machine learning models, allowing developers to incorporate their own trained models or leverage pre-trained models provided by MediaPipe. This integration enhances the capabilities of applications, enabling tasks such as object detection and facial recognition.

The framework supports various platforms, including Android, iOS, and desktop environments, making it adaptable for different devices. It prioritizes efficiency, capable of running in real-time on a range of devices, including mobile phones and edge devices.

MediaPipe also provides visualization tools to aid developers in understanding the output of each pipeline component, facilitating debugging and optimization. Overall, MediaPipe simplifies the development of real-time computer vision applications by offering a flexible, efficient framework with pre-built components for common tasks.

2) RIPPER algorithm: The RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm is a widely used rule induction algorithm for classification tasks. [1] It is known for its efficiency in handling large datasets and its ability to generate accurate and interpretable rules. The RIPPER algorithm works in three main steps -

a) Rule Growing: The algorithm starts with an empty rule and iteratively adds conditions to the rule until it perfectly classifies a subset of the data. At each iteration, the algorithm selects the attribute that maximizes the information gain, which measures the reduction in uncertainty about the class label after considering the attribute value. The process stops when the rule starts covering negative examples (examples that do not belong to the target class).

b) Rule Pruning: : The algorithm then prunes the rule to remove any redundant or unnecessary conditions. Pruning helps to prevent overfitting and improve the generalization ability of the rules. RIPPER uses a technique called incrementally reduced error pruning (IREP) to prune the rule. IREP works by repeatedly splitting the rule into two subrules and evaluating each subrule's performance on a separate validation set. The subrule with the lower error rate is kept, and the other subrule is discarded. This process is repeated until no further improvements can be made.

c) Rule Ordering: : Finally, the algorithm orders the rules in a descending order of their accuracy. This ordering ensures that the most accurate rules are applied first, which can improve the overall performance of the classifier.

The RIPPER algorithm has several advantages over other rule induction algorithms. RIPPER is able to handle large datasets efficiently due to its incremental approach to rule growing and pruning. RIPPER is relatively robust to noise in the data due to its use of a validation set for pruning. The rules generated by RIPPER are relatively easy to interpret, which can be helpful for understanding the underlying patterns in the data.

B. Training Phase

1) Facial Feature Extraction using Mediapipe: The initial step in the training phase involves the utilization of the Mediapipe framework for facial feature extraction. This tool is chosen for its efficiency in capturing a wide array of facial landmarks and features, providing a holistic representation of the facial expressions present in the training images. The detailed feature vectors obtained serve as the foundation for subsequent training steps.

The selection of the Mediapipe framework for facial feature extraction is rooted in its versatility and accuracy in capturing detailed facial expressions. Its ability to provide a rich set of facial landmarks ensures a comprehensive representation of the emotional cues present in the images. However, it's essential to acknowledge that the choice of facial feature extraction tool is a critical aspect that may impact the overall performance of the system. Alternative tools and methodologies could be explored in future iterations to assess their efficacy in enhancing feature extraction accuracy.

2) Label Encoding: Post-feature extraction, the emotions associated with the training images undergo a meticulous label encoding process. The one-hot encoding technique is employed, creating distinct representations for each emotion category. The pivotal role of Label Encoding is elucidated as a crucial preprocessing step to enhance the applicability of the RIPPER algorithm. The underlying principle of Label Encoding involves the transformation of emotion labels into a binary representation — 0 and 1 — corresponding to the negative and positive classes, respectively. This encoding paradigm is meticulously applied to each emotion label within the dataset, wherein the label associated with a specific emotion is set to 1, while labels for all other emotions are systematically assigned 0.

Following this encoding process, the dataset is subjected to training, enabling the RIPPER algorithm to derive intricate rules tailored to the nuanced characteristics of the targeted emotion. This iterative procedure is systematically repeated for each distinct emotion, thereby engendering the extraction of discerning rules that govern the classification of data points across various emotional states. Through the judicious application of Label Encoding, the RIPPER algorithm is adeptly configured to discern and classify diverse emotional states, thereby establishing a robust foundation for the systematic extraction of emotion-specific rules. This step establishes a clear association between the extracted facial features and their corresponding emotional labels, laying the groundwork for supervised learning.

3) RIPPER Algorithm: The training data, consisting of encoded facial features and associated emotions, is then subjected to the RIPPER (Repeated Incremental Pruning to

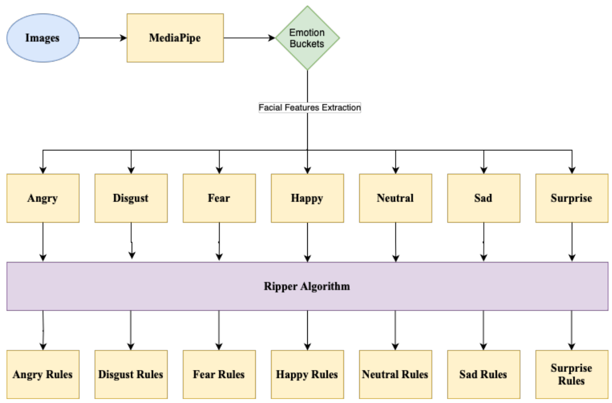

Fig. 3. Flowchart of the training phase of the RIPPER Classifier

Produce Error Reduction) algorithm. This rule-based learning algorithm excels in generating comprehensible rule sets while maintaining competitive classification accuracy. The iterative nature of RIPPER allows it to adapt and refine rules based on the intricacies of the training data, establishing a robust foundation for emotion-specific classifiers. Fig. 3 illustrates the flowchart for the training phase of the Classifier using the RIPPER algorithm.

4) Rule Generation for Each Emotion: The RIPPER algorithm is applied independently for each emotion category, resulting in a set of unique and tailored rule sets. These rules encapsulate the nuanced relationships between specific facial features and the corresponding emotional expressions. The granularity achieved through this step enhances the model's ability to discern subtle variations in facial expressions associated with different emotions.

The RIPPER algorithm is chosen for its balance between interpretability and classification accuracy. The generated rule sets provide insights into the specific facial features that contribute to emotional expressions.

C. Testing Phase

1) Facial Feature Extraction using Mediapipe: The testing phase commences with the processing of new images through the Mediapipe framework. This consistent pre-processing step ensures a standardized input for the subsequent classification stages, promoting comparability between the training and testing processes.

2) Rule-based Classification: Extracted features from test images are then subjected to the rule sets generated during the training phase. This rule-based classification approach allows for a nuanced analysis, taking into account the specific patterns associated with each emotion. The interpretability of rule-based systems is a notable advantage, enabling a deeper understanding of the decision-making process. This part of the testing phase allows the image to be classified into zero, one or multiple emotions. This is due to the fact that the test image is being passed through the rules of every emotion, giving a result whether the image is classified into that particular

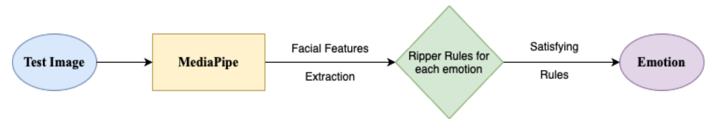

Fig. 4. Flowchart of the testing phase of the RIPPER Classifier

emotion or not. Fig. 4 illustrates the flowchart for the testing phase of the Classifier using the RIPPER algorithm.

3) Coverage Analysis: A single image may be associated with zero, one, or more emotions based on the application of multiple rule sets. Images that present challenges in straightforward classification or are identified with multiple emotions undergo a rigorous coverage analysis. The rules are re-applied to ascertain the coverage of each emotion. A coverage value is computed, representing the degree to which the image satisfies the conditions specified in each emotion's rule set. This experimental approach aims to refine multi-emotion classifications and determine the dominant emotion. It's crucial to note, however, that this step, though conceptually intriguing, exhibited limitations in preliminary experiments, indicating the necessity for further refinement and exploration in subsequent research iterations.

Algorithm 1 RIPPER Algorithm

Input: Training data D, Maximum number of rules K
Output: Set of rules R
procedure RIPPER(D, K)
    $R \leftarrow \emptyset$ ▷ Initialize the set of rules
    for $k \leftarrow 1$ to K **do**
        $r \leftarrow$ LearnRule(D, R) ▷ Learn a new rule
        if r is not null **then**
            $R \leftarrow R \cup \{r\}$ ▷ Add the new rule to the set
            $D \leftarrow D \setminus \{$(instances covered by r)$\}$ ▷ Remove covered instances
        else
            break ▷ No more rules can be learned
        end if
    end for
    return R
end procedure
procedure LEARNRULE(D, R)
    $r \leftarrow$ FindBestRule(D, R) ▷ Find the best rule for the current iteration
    if r is not null **then**
        return r
    else
        return null
    end if
end procedure
procedure FINDBESTRULE(D, R)
    ... ▷ Implementation of rule selection based on heuristics
end procedure

IV. Other Experiments

A. Decision Tree

In one of our experiments, we integrated a Decision Tree classifier from the scikit-learn library, a widely used machine learning toolkit in Python, to bolster our emotion recognition framework. The choice of this classifier was motivated by its capability to construct clear and interpretable models, essential for deciphering the complex process of emotion recognition.

The classifier, employing a tree-like decision model, was trained using facial features extracted with the MediaPipe library. In this model, each node represents a specific facial feature, while the branches indicate decision rules that lead to various emotional states. The training process involves dividing the dataset into subsets based on these features, sequentially selecting the most influential feature at each node to further split the data. This method enables the model to predict an emotion by analyzing an image's features and following a path down to a leaf node, which signifies the predicted emotion. While this approach provides a straightforward method for classifying emotions, our findings highlight some limitations. The trained decision tree model achieved an overall accuracy of **34%**, indicating challenges in capturing the nuanced and complex variations inherent in human emotions. This outcome suggests the potential need for more sophisticated or hybrid models in accurately interpreting emotional states from facial features.

1) Decision Tree Classifier Results: Table I presents the performance metrics of the Decision Tree Classifier, which was trained using the facial features dataset. These metrics include accuracy, precision, recall, and F1-score, providing a comprehensive overview of the model's performance in emotion recognition. Additionally, Fig. 5 illustrates the confusion matrix for the Decision Tree Classifier, offering a visual representation of the model's performance in terms of correctly and incorrectly classified instances across different emotional states.

TABLE I
Metrics with Decision Tree Classifier

Emotion	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Angry	30.27	30	30	30
Disgust	2.70	3	3	3
Fear	13.36	12	13	12
Happy	50.20	55	50	52
Neutral	36.74	38	37	37
Sad	26.16	24	26	25
Surprise	28.95	29	29	29

B. Apriori Algorithm

The Apriori algorithm is a classic algorithm used in data mining for the extraction of frequent itemsets and the discovery of association rules. It operates on a simple principle: if an itemset is frequent, then all of its subsets must also be frequent. This approach helps in significantly reducing the number of itemsets the algorithm needs to examine.

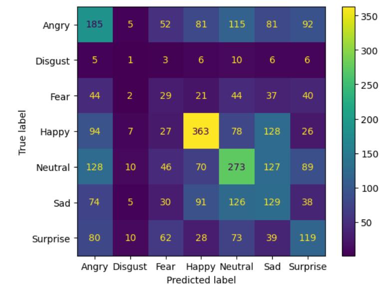

Fig. 5. Confusion Matrix for the trained Decision Tree Model

$('browInnerUp', 'mouthPucker', 'browOuterUpLeft') \implies ('Surprise')$
$('browInnerUp', 'mouthPucker', 'browOuterUpLeft', 'browOuterUpRight') \implies ('Surprise')$
$('eyeSquintLeft', 'eyeSquintRight', 'eyeLookUpLeft') \implies ('Neutral')$
$('eyeSquintLeft', 'eyeSquintRight', 'eyeLookUpLeft', 'eyeLookUpRight') \implies ('Neutral')$
$('eyeLookOutLeft', 'eyeSquintRight', 'eyeLookInRight', 'eyeSquintLeft') \implies ('Neutral')$
$('eyeLookOutRight', 'eyeSquintLeft', 'eyeSquintRight', 'eyeLookInLeft') \implies ('Neutral')$
$('browInnerUp', 'mouthPucker', 'browOuterUpLeft') \implies ('Neutral')$
$('eyeSquintLeft', 'browDownLeft', 'eyeSquintRight') \implies ('Neutral')$
$('eyeSquintLeft', 'browDownLeft', 'eyeSquintRight', 'browDownRight') \implies ('Neutral')$
$('mouthSmileRight', 'eyeSquintLeft') \implies ('Happy')$

Fig. 6. Rules generated by Apriori

Apriori works in two steps: First, it identifies the frequent individual items in the database and extends them to larger itemsets as long as those itemsets appear sufficiently often in the database. The frequency is determined by a user-specified parameter called 'support'. Second, the algorithm uses these frequent itemsets to generate association rules.

1) Rule Derivation Using Apriori Algorithm: With a scaled dataset obtained using the Min-Max Scaler, we utilized the Apriori algorithm for mining frequent item sets and deriving association rules among the facial landmarks. These rules are based on the likelihood of certain facial features co-occurring with specific emotions. We measured the strength of these rules using metrics such as support, confidence, and lift. Figure 6 shows the association rules that were generated after the facial features were passed through the Apriori algorithm.

2) Refinement of Association Rules: Initially, the derived rules lacked direct association with emotion labels, presenting relationships only between various facial features. To address this, we integrated emotion labels into our dataset and employed the MLxtend library's built-in Apriori algorithm. This enabled us to derive rules that associate specific combinations of facial features with particular emotions, along with their corresponding support and confidence values.

3) Emotion Classification Strategy: Our classification strategy involved comparing the facial features of test images against the derived rules. An image's emotion was inferred based on the highest-confidence rule that its features matched. In cases where features matched multiple rules, the rule with the highest confidence value was selected to determine the emotion.

Algorithm 2 Apriori algorithm

begin
  $L_1 \leftarrow Frequent1 - itemset$
  $k \leftarrow 2$
  while $L_{k-1} \neq \phi$ **do**
    $Temp \leftarrow candidateItemSet(L_{k-1})$
    $C_k \leftarrow frequencyOfItemSet(Temp)$
    $L_k \leftarrow compareISetW/MinSupport(C_k, minsup)$
    $k \leftarrow k + 1$
  end while
  return L
end

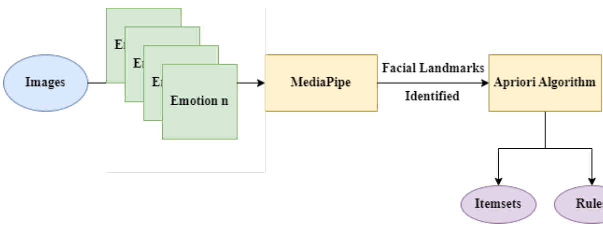

Fig. 7. Flowchart of Emotion Recognition System using Apriori

4) Algorithm Implementation and Training Data: The Apriori algorithm's implementation involved the following steps - Identify frequent one-item sets, Iteratively explore larger item sets based on the frequency of occurrence, Compare these sets against a minimum support threshold.

5) Challenges in Precision and Recall: In the implementation phase, we encountered challenges with precision and recall metrics. Despite accurately classifying certain emotions, the system showed limitations in consistently identifying all emotional states, indicating areas for further refinement and optimization of our methodology.

C. Combining Apriori Rules and RIPPER Rules

In an innovative approach to enhance our emotion recognition system, we experimented with the integration of association rules derived from both the Apriori and the RIPPER algorithms. The Apriori algorithm, as previously described, excels in identifying frequent itemsets and generating rules based on the likelihood of occurrence of certain facial features together. Meanwhile, the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm, another rule-based learning method, is known for its efficiency in creating compact and interpretable rule sets for classification.

Our hypothesis was that the combination of these two methodologies could potentially leverage the strengths of both - the comprehensive rule generation of Apriori and the efficient, concise rule creation of RIPPER. We aimed to create a more robust rule set that could better handle the complexities and subtleties of emotion recognition from facial features.

To implement this, we first generated a set of rules using the Apriori algorithm, which provided us with associations between various facial features and emotional states. Following this, we applied the RIPPER algorithm to the same dataset to derive a different set of rules, which were expected to be more concise and focused.

The next step was the integration of these two sets of rules. This process involved a careful examination and combination of the rules from both algorithms, aiming to create a hybrid set that would ideally be more effective than the individual sets in classifying emotions.

However, upon evaluating this combined rule set against our metrics, which included accuracy, precision, and recall, we did not observe a significant improvement in the performance of our emotion recognition system. The combined rule set did not demonstrate enhanced predictive capabilities compared to the rules generated individually by either the Apriori or the RIPPER algorithm.

Given these results, we concluded that the integration of Apriori and RIPPER rules, while theoretically promising, did not yield practical benefits in our specific application. The lack of improvement led us to decide against pursuing this combined approach further in our project. This outcome highlights the challenges in emotion recognition tasks and suggests that the complexity of human emotions may not be fully captured by simply merging different rule-based methods.

1) Results from combining rules from Apriori and RIPPER Algorithms: Table II presents the evaluation metrics resulting from the application of combined rules derived from the Apriori and RIPPER algorithms. This table includes various performance indicators such as accuracy, precision, and recall, which provide insights into the effectiveness of integrating these two rule-based methods in emotion recognition.

TABLE II
METRICS FOR CLASSIFIER WITH COMBINED RULES FROM RIPPER AND APRIORI ALGORITHMS

Emotion	Accuracy (%)	Precision (%)	Recall (%)
Angry	82.11	44.44	2.08
Disgust	98.88	-	-
Fear	92.95	-	-
Happy	83.87	70.68	50
Neutral	77.38	12.5	0.42
Sad	81.92	25	1.58
Surprise	87.48	46.15	13.63

V. TRAINING DATA

We train our system on Dynamic Facial Expression in-the-Wild (DFEW) [5]. It is a large-scale facial expression database with challenging video clips taken from movies. Clips in the DFEW database are of various challenging interferences, such as extreme illumination, occlusions, and capricious pose changes. Based on the crowdsourcing annotations, they hired 12 expert annotators, and each clip has been independently labeled ten times by them. DFEW database has enormous diversities, large quantities, and rich annotations, including: 16372 number of very challenging video clips from movies, a 7-dimensional expression distribution vector for each video clip, single-labeled expression annotation for classic seven discrete emotions and baseline classifier outputs based on single-

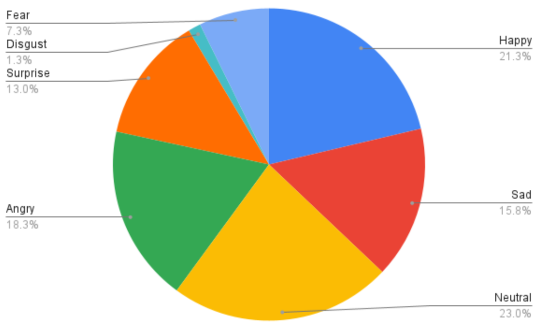

Fig. 8. Distribution of Images based on their emotions

```
[[mouthSmileLeft=>0.46 ^ mouthUpperUpLeft=>0.1 ^ mouthPressRight=0.065-0.12 ^
cheekSquintRight=>1.9e-06] V
[mouthSmileLeft=>0.46 ^ mouthUpperUpLeft=>0.1] V
[mouthSmileRight=0.07-0.45 ^ noseSneerLeft=2.8e-06-4e-06 ^ browInnerUp=0.34-0.5] V
[mouthSmileRight=0.07-0.45 ^ jawOpen=<0.0011 ^ eyeWideRight=0.012-0.019] V
[mouthSmileLeft=0.074-0.46 ^ mouthFrownLeft=<5.1e-05] V
[mouthSmileLeft=>0.46 ^ mouthPucker=0.0007-0.002 ^ mouthRollLower=0.018-0.029] V
[mouthSmileRight=0.07-0.45 ^ jawOpen=<0.0011] V
[mouthSmileRight=>0.45 ^ eyeWideLeft=0.0084-0.012 ^ mouthUpperUpLeft=0.016-0.1] V
[mouthSmileRight=>0.45 ^ mouthFrownLeft=0.00019-0.00046 ^ mouthPressLeft=0.051-0.095]
........
```

Fig. 9. Rules generated by using RIPPER for the emotion "Happy"

labeled annotation. Fig. 8 shows the distribution of classes or emotions in the dataset. As a whole, we are considering 10782 images from the dataset which consists of 2298 happy images, 1699 sad images, 2479 neutral images, 1975 angry images, 1405 surprised images, 140 disgust images and 786 fear images.

VI. Results

The adoption of the RIPPER algorithm, coupled with strategic dataset optimization, substantially elevated the system's ability to discern emotions with higher precision and recall. The exclusion of certain emotions from the dataset may seem counterintuitive at first glance, as it reduces the classifier's ability to recognize a broader spectrum of emotional states. However, this was a calculated trade-off to improve the system's reliability and confidence when detecting the more frequently represented emotions. Figure 9 shows the subset of rules that were generated using the RIPPER algorithm for the "Happy" emotion. Table III shows the metrics of the RIPPER classifier for each emotion when we normalize the facial features dataset using Min-Max Scaling while Table IV shows the metrics without applying any normalizing techniques to the facial features dataset. Tables V and VI show the improvement in the Classifier's precision and recall when we consider only the frequently represented emotions.

The transition from the Apriori algorithm to the RIPPER algorithm has yielded a dramatic enhancement in the performance metrics of our emotion recognition model. The accuracy scores have seen a significant upsurge, underscoring the limitations of the previous model's ability to generalize across the spectrum of emotions. By omitting the min-max scaling, our model's recall improved markedly, indicating a

TABLE III
Metrics with Min-Max Scaling

Emotion	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Angry	81.46	37.93	2.79	5.2
Disgust	98.89	-	-	-
Fear	93.23	16.66	0.70	1.34
Happy	83.17	68.79	46.61	55.57
Neutral	77.38	67.74	4.21	7.93
Sad	83.91	39.13	2.63	4.82
Surprise	87.02	41.30	6.98	11.94

TABLE IV
Metrics without Min-Max Scaling

Emotion	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Angry	81.83	51.52	4.33	7.99
Disgust	98.79	16.67	4.55	7.15
Fear	93.32	-	-	-
Happy	83.40	69.85	46.61	55.91
Neutral	76.63	45.45	5.01	9.02
Sad	84.28	63.64	2.05	3.97
Surprise	87.34	49.02	9.19	15.48

TABLE V
Metrics for Three Emotions with Min-Max Scaling

Emotion	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Angry	53.23	64.12	5.68	10.44
Happy	79.02	74.33	58.45	65.44
Neutral	55.96	44.12	19.72	27.26

TABLE VI
Metrics for Three Emotions without Min-Max Scaling

Emotion	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Angry	71.13	70.69	9.90	17.36
Happy	81.13	76.10	62.25	68.48
Neutral	66.77	64.24	19.72	30.17

heightened sensitivity to the correct identification of emotional states. This targeted approach not only improved the model's discriminative power but also its overall reliability in classifying emotions accurately.

However, it's important to note that while our model has good accuracy, the recall values for specific emotions, though not explicitly mentioned in the table, reveal areas for improvement. Low recall suggests that our model may have difficulty correctly identifying certain emotions, leading to false negatives.

Table VII provides a comparison of accuracy scores among various models, including C3D, P3D, R3D18, 3D Resnet18, I3D-RGB, VGG11+LSTM, Resnet18+LSTM, and our model.

An insightful observation from our experiments is the sensitivity of emotion recognition to the order in which the Coverage Analysis is applied. The order in which rules are applied to features extracted from an image emerges as a significant hyperparameter, indicating that the sequencing of emotion rules may significantly influence the overall model

TABLE VII
COMPARISON WITH BASELINE MODELS

Model	**Happy**	**Sad**	**Neutral**	**Angry**	**Surprise**	**Disgust**	**Fear**
C3D	75.17	39.49	55.11	62.49	45.00	1.38	20.51
P3D	74.85	43.40	54.18	60.42	50.99	0.69	23.28
R3D18	79.67	39.07	57.66	50.39	48.26	3.45	21.06
3D Resnet18	73.13	48.26	50.51	64.75	50.10	0.00	26.39
I3D-RGB	78.61	44.19	56.69	55.87	45.88	2.07	20.51
VGG11+LSTM	76.89	37.65	58.04	60.70	43.70	0.00	19.73
Resnet18+LSTM	78.00	40.65	53.77	56.83	45.00	4.14	21.62
Our Model	50.87	7.31	40.08	6.63	23.77	4.54	6.87

performance. This finding underscores the complexity of emotion recognition and opens avenues for future investigations aimed at refining and optimizing the order in which emotion rules are applied. Consequently, our ongoing research agenda includes an in-depth exploration of this hyperparameter, with the aim of enhancing the overall performance and robustness of emotion recognition models. We anticipate that further refinements in the model architecture, coupled with a meticulous investigation into the optimal order of processing emotions, will contribute to the advancement of emotion recognition technology and its application in real-world scenarios.

VII. CONCLUSION AND FUTURE SCOPE

This research represents a significant advancement in the field of facial emotion recognition by integrating perceptual reasoning and symbol grounding within computational frameworks. Our initial methodology, which combined Google's MediaPipe library for feature extraction with the Apriori algorithm for association rule learning, laid a strong foundation for symbolizing emotional states from facial features. Although the accuracy for detecting happiness was moderately successful at **43.75%**, it indicated the complexity of emotion recognition and the need for further enhancements.

The incorporation of the RIPPER algorithm to extract classification rules has markedly improved our system's performance, specifically by refining the precision of emotion classification. The transition from the Apriori to the RIPPER algorithm has also addressed some of the scalability concerns associated with large datasets, leading to a more efficient rule-based recognition system.

The future advancements of this research are multifaceted, primarily focusing on enhancing the robustness and applicability of the emotion recognition system. Key areas for development include algorithm refinement and exploration of alternative rule-learning algorithms to improve precision and recall across a broader spectrum of emotions. Expanding the range of facial features and investigating other feature extraction methods could significantly augment the model's capability to discern subtle emotional expressions. Additionally, integrating our approach with other machine learning techniques, such as deep learning, could result in a hybrid model that harnesses the strengths of different methodologies for a more comprehensive emotion recognition system. Utilizing more diverse and extensive datasets, including those representing various demographic and cultural backgrounds, is essential to enhance the model's generalizability and inclusiveness. Collaborations with experts in psychology, neuroscience, and human-computer interaction can provide deeper insights into human emotions, leading to more sophisticated and nuanced models.This comprehensive future scope sets the stage for transforming the emotion recognition field, making it more accurate, interpretable, and aligned with human-centric applications.

REFERENCES

[1] William W. Cohen. Fast effective rule induction. In Armand Prieditis and Stuart Russell, editors, *Machine Learning Proceedings 1995*, pages 115–123. Morgan Kaufmann, San Francisco (CA), 1995.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.

[3] Spiros V. Ioannou, Amaryllis T. Raouzaiou, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzis, and Stefanos D. Kollias. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks*, 18(4):423–435, 2005. Emotion and Brain.

[4] Charvi Jain, Kshitij Sawant, Mohammed Rehman, and Rajesh Kumar. Emotion detection and characterization using facial features. In *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–6, 2018.

[5] Xingxun Jiang, Yuan Zong, Wenming Zheng, Chuangao Tang, Wanchuang Xia, Cheng Lu, and Jiateng Liu. Dfew: A large-scale database for recognizing dynamic facial expressions in the wild. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, page 2881–2889, New York, NY, USA, 2020. Association for Computing Machinery.

[6] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for perceiving and processing reality. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*, 2019.

[7] Chung-Hsien Wu, Ze-Jing Chuang, and Yu-Chung Lin. Emotion recognition from text using semantic labels and separable mixture models. *ACM Transactions on Asian Language Information Processing*, 5(2):165–183, jun 2006.